Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in
Lingue Moderne per la Comunicazione e la Cooperazione Internazionale
Classe LM-38

Tesi di Laurea

Italian and German students' use of the verb *get*: a learner corpus analysis

Relatrice
Prof. Katherine Ackerley

Laureanda
Jessica Malosso
n° matr.2007039 / LMLCC

Anno Accademico 2022 / 2023

Table of Contents

INTRODUCTION

In this study I aim to investigate how Italian and German learners of English use the verb *get* through the analysis of corpora: in particular, I focus on which meanings of the verb are the most frequent, which types of errors learners make when producing it and in which aspects the two categories are similar to each other or present peculiar differences due to their native tongue.

To start with, this thesis is divided in four chapters: in the first one I will give a general overview of what corpora are and of the different types in existence. In fact corpora, which are collections of texts, can be of different sizes, of different typology and can contain different forms of language: it can be spoken as well as written, it can be of a determinate register or period of time, and it can be produced by natives as well as by learners of different L1s.

I will focus especially on the advantages that L1 corpora, which are collections of texts produced by L1 speakers of a particular language, can offer in the education environment starting from the advent of technology: especially through frequency lists and concordance lines, which can be extracted through specific software, it is possible in fact to quickly observe determinate features of the language present in the corpora analysed and see the contexts in which they are produced, which can be helpful in the creation of dictionaries and coursebooks; moreover, tasks can also be created starting from these, as the 'fill in the blanks' type where words from concordances can be taken out, and studies of many different types can be carried out. This inductive method of using corpora in the context of language learning, called data-driven learning, can be particularly useful in the classroom environment and can represent an interesting and motivational activity for students and teachers, who can both act as researchers; learning about how corpora can be benefited from both perspectives has been particularly interesting for me since this past year I experienced both positions, by attending my last year of university in Modern Languages and by starting teaching English at some middle schools.

In the last part of the chapter I will then report some studies by authors who have investigated both how corpora can be used not only in a university context but also in the first and intermediate years of learners' education (Sealey and Thompson 2004, Chujo, Utiyama and Miura 2006, Geist and Hahn 2012), and the differences between the paper-

based and the computer-based approaches of data-driven learning (Chujo and Oghigian 2012, Boulton 2012).

As my study will be based on two components of a learner corpus, in the second chapter I will provide more detailed information about learner corpora only: these are collections of texts produced by L2 learners of a determinate language and for this reason they can be used and exploited by learners and teachers in different ways as compared to native corpora. While L1 corpora offer the authentic language and its correct uses, L2 corpora can show the mistakes and inappropriate uses which learners of a language make, and thus what is more difficult for them to understand and reproduce. These features can be included as well in dictionaries and coursebooks, through 'help boxes' aimed to show the mostly common misuses of words and structures, and also be of benefit to teachers, who can create exercises to explain any particular topic at school which involve, for example, error identification or correction. Materials can be created out of the mistakes of the students themselves as well and this can represent a benefit for them too, as they start seeing errors in a more positive perspective.

After their pedagogical uses, I will focus on previous research based on learner corpora, which regard especially learners of the English language and the International Corpus of Learner English (ICLE), the corpus which I will use too for my study; these studies often take into consideration more than one L2, in order to see and compare determinate peculiarities of specific learners of English. In particular, two sections will be dedicated to Italian and German learners only, in order to already form an idea about how these two categories approach some aspects of the language: these include collocations (Nesselhauf 2005, Mukherjee 2009) and cleft sentences (Prat Zagrebelsky 2004, Callies 2009, Castello 2017) among others. A final section will then be about studies on spoken learner corpora, as they represent an interesting area of study in the context of fluency and of devices used in the oral production of English.

In the third chapter I will introduce the topic of my research, which is the verb *get*: I will provide the main meanings of the verb, as well as the most frequent phrasal verbs and idioms which it can form and other previous studies about them.

Since it represents one of the main polysemous and high-frequency verbs in English, this verb can have many uses: it can express possession when followed by a noun, both in the active sense of acquiring (*get money*) and in the passive one of receiving (*get help*); it can

express movement when linked to other prepositions (*get to school*); it can then be linked to adjectives, and in this sense mean a change of state, in its inchoative use (*get interested*), or form a passive structure, which often denotes more focus on the action and a more negative connotation as compared to the *be* passive (*get killed*); it can also be meant in a metaphorical way (*get a feeling*) and through its past form *got*, it can form a modal with a sense of obligation (*have got to go*). Two other functions of *get* which are formed by more arguments are the causation one (*he got me a present*) and the ingressive one, which expresses the beginning of a state (*they eventually got to like me*).

A more detailed overview of these main meanings of the verb will be followed by a list of the main phrasal verbs and idioms which *get* can form: these structures are formed by a verb plus a particle and their main peculiarity is that they both have a meaning not deducible by its single components. Phrasal verbs often have more than one meaning as well, and for these reasons they represent one particularly difficult area for learners, as high-frequency verbs themselves: the studies provided in the final section of this chapter will take in consideration especially potential factors that influence learners' production or lack of them, as their native language (Dagut and Laufer 1985, Hulstijn and Marchena 1989), their learning context (Becker 2014) and activities such as reading or listening to music (Schmitt and Redwood 2011) regarding phrasal verbs, and textbooks' design (Gouverneur 2008) regarding high-frequency verbs. One final study (Coto-Villalibre 2016) will then be about *get* only: since it considers Hong Kong students, it will thus be possible to see first the results obtained from learners with another L1, which will be compared to those found in my own study.

The fourth and last chapter will be my personal research of how the verb *get* is used by German and Italian learners: I chose these two categories since they are my native language and the other foreign language I studied in my degree along with English, and the verb *get* as it is among the most frequent verbs in the English language and therefore common in both written and spoken register. Since as mentioned above it can have multiple meanings and it forms many phrasal verbs and idioms, it constitutes a rather challenging learning topic for students, and for this reason the main questions I was interested to investigate were if the two categories produce a variety of the uses of the verb illustrated, if they present similarities, especially in the type of mistakes they tend to make, and if there are considerable differences when comparing their production with that

of a L1 corpus.

For the data, I will analyse the Italian and German sub-components of the ICLE, particularly through a software called *AntConc* in order to retrieve all examples of the verb in the form of concordance lines. I will provide a section for each meaning found in the corpora and a final one about mistakes, linked to tables with their frequency in terms of number of hits and of normalized frequency scores too, which take into account the number of total tokens of each corpus and provide therefore more detailed figures.

After comparing the results of the two categories of L2 learners, in the last part of the chapter I will take into consideration also the L1 corpus LOCNESS, made up of essays produced by British and American students of English, first analyzing how these L1 students use the same verb, and then comparing first native and learner corpus' results and eventually each single category of them, in order to obtain a more complete vision of the verb and also in order to suggest suitable tasks that could be useful for students and teachers.

This topic was partly inspired by a course I attended at TU Dresden during my Erasmus program in Germany called Error Analysis: every week, the Professor provided us with a set of ten sentences produced in the English language by native speakers of German (retrieved during conversation classes and presentations of English Language courses at Southampton English Language Centre), and our task was to identify and explain the mistakes present in them; as I found this course particularly interesting but also fun, and since I realized that I learned a lot more about German but also about English itself, I wanted to do research of this type in my thesis, taking into account learners of my native tongue as well.

CHAPTER 1

1. CORPORA AND THEIR USE IN LANGUAGE TEACHING AND LEARNING

In this first chapter, I provide a description of what corpora are and how the latest innovations in technology have changed the way they can be used, in particular by students and teachers while studying and teaching languages.

1.1 Introduction to corpora

These last years of technology evolution have completely changed the way languages can be seen, studied and taught, particularly through corpora. A corpus can be defined as "a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study" (Hunston 2002: 2). The main purpose of a corpus is thus to study how languages work and which structures are typical of them: being variability one of the most distinctive characteristics of languages, the possibility to gather many different authentic examples from different sources through a large corpus allows their regularities to be identifiable.

What distinguishes a corpus from a text, according to Sinclair (2004), is the methodology: while a text can be read as a whole from the beginning to the end, the content of a corpus cannot be read directly but through particular tools that allow new perspectives on words and patterns to be visible and analysed.

Corpora can vary from various aspects: they can be of different sizes, they can contain whole documents or just samples; regarding the content, they can be a collection of written as well as spoken texts and these can be informal or formal, depending on the aim of the research. In particular, eight specific typologies of corpora can be identified and described (Hunston 2002): general (or reference) corpora, which include texts of many types and thus contain the standard vocabulary of a language and are important for studying lexical semantics; specialised corpora, which contain texts of a particular type

and therefore aim to be representative of that type of language; comparable corpora, which contain two or more corpora in different languages or in different varieties of a language; parallel corpora, which also contain two or more corpora in different languages but each containing texts that have been translated from one language into the other, making therefore possible to find a lot of translation equivalents; monitor corpora, which are designed to keep track of the changes in a language and are constantly updated; historical corpora, which contain texts from different periods of time; learner corpora, which include texts produced by learners of a language, and finally pedagogic corpora, which consist of all the language a learner has been exposed to, meaning all the words and expressions encountered in all their various contexts. It is thus possible to refer to corpora as collections of data about many aspects of language, that vary according to the specific type and aim of the corpus.

Before the advent of computers, this collection of materials was manual and therefore time-consuming, thus as soon as the potential of new technologies was discovered, linguists began to take advantage of it: new research methods have been developed, large amounts of language data could be viewed and stored electronically in a way that was thought to be impossible until recently and this "ability to [...] access and retrieve this data through a software interface has paved the way for the emergence of corpus linguistics" (Breyer 2011: 1). What is referred to as corpus linguistics is an approach that "studies languages on the basis of discourse" (Teubert 2004: 100) and thus regards a corpus as a representation of a discourse, defining such as the totality of texts and interactions produced, over centuries, by the members of the language's discourse community; it therefore "sees language as a social phenomenon" (Teubert 2004: 97), since discourse is produced by the people who speak that language and who decide which words to use and what their meaning is in a specific period of time. As the author argues, the meaning of words is not a fixed construct like any grammatical rule, but it is vague and for this reason many words are considered ambiguous. A difference which needs to be pointed out is therefore the one between words and units of meaning, the latter being words embedded in a context, and consequently unambiguous. Both language and meanings, being a product of the discourse community as mentioned above, are subject to changes; since words are just symbols, people can either introduce new ones or change the meaning of an already existing word: Teubert and Čermáková (2004) mention as an

example the word ‘wicked’, that from an original significance of ‘mean’ nowadays is used for something good, above all in slang. Changes can also be due to historical circumstances, as shows the example that the authors make of the word ‘globalization’: a look at an article from 1983 shows that it was only mentioned as a derived noun from ‘global’ and appeared in only nine sentences, while after the 90’s and the popularity that the word gained, it became established in the English language.
In conclusion, considering that new words and meanings are coined constantly, especially through the Internet and the significant use of social networks in our modern societies, creating a corpus is a continual challenge that can be carried on.

1.2 Methods of analysis

The most useful processing methods through which researchers actually extract information from computer corpora are frequency lists, phraseology and collocations.
As for frequency, it is possible to distinguish between ‘raw frequency’, which is the count of how many instances of a linguistic phenomenon occur in a corpus or text, ‘normalized frequency’, which expresses frequency relative to a standard yardstick and ‘ordinal frequency’, which is the frequency of a linguistic phenomenon compared with the frequencies of other phenomena (Leech 2011): therefore if we take into account the standard definition of frequency lists as “lists of all the words that occur in a corpus complemented by the number of occurrence of each individual item” (Breyer 2011: 35), we are referring to the definition of ‘ordinal frequency’. As Leech further explains, word frequency was considered very important in the beginning of the 20th century in language teaching, as was the postulate “more frequent = more important to learn”; then in the second half of the century this concept was rejected by Chomsky, who stated that frequency had no linguistic relevance and probability had nothing to do with grammar; only with the advent of computers and with more corpus-based studies, the perspective on frequency shifted again in its favor. The postulate mentioned above is still relevant, as it is based on the assumption that if a word or an expression appears several times in the discourse of communities, it means that it is entrenched in that language and it is thus essential for learners to know how to use it.

According to the specific type of corpus that is being analysed, many observations can be made (e.g. lexical words will be more frequent than grammar ones in specialised corpora and will therefore be referred to as 'keywords'), also through comparisons between more corpora: this is important above all in the beginning of a research, as it will be an initial visualisation of the most prominent items in the texts selected.

The information included in the frequency lists can be illustrated in alphabetical or in frequency order, both also in ascending or descending order (Sinclair 1991).

It is important to note, however, that if a word is particularly frequent it does not have to mean that it is also significantly used: this represents an initial step and further analysis is often needed. At the same time, words which are not very frequent can also be important if they have a cultural resonance: Hunston (2002) refers to the words *death* and *adventure*, which generally were not found together with a positive correlation, however their occurrence in the books of '*Peter Pan*' and in the '*Harry Potter*' saga, which are among the most famous and read all over the world, changed this perspective. A look through corpora can thus help in investigating the variety of aspects and features of the language and the ways through which it can develop and change.

Phraseology, which can be defined as the study of words and phrases organised in set expressions, can be analysed through concordance lines, which "bring together many instances of use of a word or phrase, allowing the user to observe regularities in use that tend to remain unobserved when the same words or phrases are met in their normal contexts" (Hunston 2002: 9). This process is feasible through concordance software, that make these patterns visible by rearranging the texts and grouping specific words according to the research in question: the most common form in which concordances are produced is through KWIC (key word in context) lists, which "present every retrieved instance of the search string, also known as the node word, in a centred column with the context displayed on the left and on the right side of that column" (Breyer 2011: 31), as shown in Figure 1.

Nr	Comment	Left context	KWIC	Right context
31		These are war-games. You're	important	because you're sitting on Cassidy's files; but for the ...
32		... a conservative selection). Drusilla Modjeska's	important	book Exiles at Home (Angus and Robertson,...
33		...-related diseases had elapsed. Thirdly, and most	important	, both Drew's estimates of tobacco-induced deaths ...
34		... which is based on eights, there is an immensely	important	boy called "The Ninth Man In The Monarch." I have ...
35		... southern Africa question. He was a member of the	important	Brandt Commission which studied questions ...
36		... of the area, especially the Fishermen Islands, are	important	breeding grounds for nine species of sea birds and ...
37		... is sad that society pigeonholes motherhood as an	important	but low-status job because the going rate for ...
38		... or by bringing more land into production. Also	important	(but perhaps more painful) has been the exit of ...
39		... of manufacturing in Queen+sland and the lack of an	important	capital city there has never really been the ...
40		... rarely in practice. In 1951 Enderby studied the more	important	case of a dilute suspension of weakly-charged ...
41		... for restraint. He also served notice on industry that	important	changes had to be made. The Government ought to ...
42		<h>3.1 INTRODUCTION </h> Eight	important	coal rich areas are discussed in this report. They are ...
43		... operating in India. <h> ZINCALUME </h> The most	important	coating development in three decades was inspired ...
44		... in the walled garden. Where no doubt a lot more	important	concerns had begun. Mr Gladstone walking the paths.
45		... of the physics of how molecules interact, with	important	consequences for the chemistry of molecular ...
46		... While an appropriate balance of State members is an	important	consideration in the selection of assess+ment panels,...
47		... the party who received it before marriage, that is an	important	consideration. (f) The strength of a contribution ...
48		... Carol leaves Lindsay. The film also looks at other	important	contempo+rary relationship issues: managing as a ...
49		... competition faced by local manufacturers, two	important	determinants of the extent to which an industry is ...
50		... and a streamlining of practices and organisation. An	important	development in 1985-86 has been a restructuring of ...

Figure 1: KWIC display (My Concordancer; Corpus: ACE) Breyer 2011: 33.

In addition to single word forms, also series of words can be searched, for example if the interest of the study is to look at adjectives that can be found in a given sequence, as in the so-called 'frames', where the first and the last words are fixed but the middle one is not (e.g. be … to: *be good to*, *be able to*, etc); furthermore, through symbols which have an open meaning denominated wildcards (Tribble & Jones 1997), it is possible to look for a word and also the forms associated with it: the most common is the asterisk (*), that can be used in researching the behaviour of similar words (for example, a research about English adverbs can begin by writing **ly* in the concordance software, since, being the most common form in the English grammar, it will provide many examples). This paradigmatic format is extremely useful since learning the vocabulary of a language does not mean to acquire just words, but also how these words are connected to one another: providing the context for each word, it shows the various meanings which that word can have according to the other words with which it is linked, representing "an intermediate position between the highly organized, graded and idealized language of the typical coursebook, and the potentially confusing but far richer and more revealing 'full flood' of authentic communication" (Johns 1986 in Braun 2010: 79). In this way, "learning the phraseology associated with a word becomes as much part of learning a word as its pronunciation or part of speech" (Hunston 2009: 142).

Concordances often produce collocations, which are frequent combinations of two or more words. In further detail, a collocation is made up of a node word and its collocates in a given span, that can be a number of words at its left or right. The node word can

either be a word form or a lemma, where the latter comprehends all the variants of a word form, like plural forms or different verb tenses (for example, GO constitutes a lemma, while *goes* or *went* are word forms). Teubert (2004) emphasises that a collocation represents one lexical item, even if it is formed by more than one word, and that its peculiarity is that it cannot be translated by translating the single components, since it provides a different meaning. Collocations illustrate in fact the so-called idiom principle, which states that words do not occur in a random order in a text, but that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments" (Sinclair 1991: 110). These phrases present some features, such as that they can allow internal variation in terms of lexis, syntax and word order; moreover, they often tend to co-occur with other words in strong collocation, with certain grammatical choices or in a certain semantic environment.

A collocation can be lexical (e.g. *spend time*), grammatical (e.g. *for example*) as well as formed by a lexical word and its grammatical context, also called colligation (e.g. *fall in love with someone*) and forms a linguistic relation that is called semantic prosody, which refers to the connotation that a word takes from the environment in which it is typically found: this can either be positive or negative, as explained through the examples mentioned by Stubbs (1995: 247) of CAUSE, which is often linked with negative words like *commotion*, *harm*, or *stress* and of PROVIDE that, on the other hand, has more positive collocates like *care*, *help*, or *shelter*.

As well as for single words, the frequency of collocations alone does not indicate their importance and there are often more factors to be considered: Stubbs (1995: 252) makes a pertinent example regarding the phrase *quintessentially English*: he affirms that it occurs pretty rarely in the corpus examined, but in this case the fact that the word *quintessentially* alone is rare and that a third of its use is with the word *English* is relevant for considering this collocation significant. Therefore we can assume that also the independent frequency of the node words or the collocates in the corpora is meaningful, as well as the size and the topic of the corpus itself.

On a final note, the main requisite for the study of phraseology and collocations, as Sinclair (1991) stresses, is the presence of large amounts of texts: only in this way it is possible to isolate and identify patterns which are actually recurrent and not just transitory.

1.3 Application of corpora

As Aston (1995: 261) affirms, "corpora constitute resources which, placed in the hands of teachers and learners who are aware of their potentials and limits, can significantly enrich the pedagogic environment": corpora can in fact be used in many ways as an aid for teaching and learning, especially in the creation of dictionaries and grammar textbooks. Since corpora have existed, the production of dictionaries and grammar textbooks has changed in such a way that it is impossible now to think of these as not influenced by them. As noted by Boulton (2010: 17), "[…] most new materials from major publishers today claim to be corpus-based to some extent, as do more and more internationally-recognised language tests", meaning that even if students do not know what corpora are, they still use them indirectly. Studies on frequency, on collocations, phraseology and variation (the latter in particular regarding registers, time periods and social groups) are essential for students to get a hang of the authentic language used by native people, and for this reason they are accounted the main resources used while studying: according to this, "the omission of a word is just as significant as its inclusion" (Sinclair 1991: 38), as it is a reflection of its real usage. Therefore dictionaries, whose aim is to illustrate the meaning of words separately rather than comparatively (Hunston 2002), now tend to give more detailed information about their various uses also in relation with other terms: current monolingual learner's dictionaries in particular provide more information about frequency by dividing the different senses and displaying the most frequent uses first, they include collocations as well as help boxes for close synonyms and confusing words, and also special entries for common errors or problems extracted from learner corpora (Boulton & De Cock 2017). These additions, together with actual sentences taken from corpora which function as examples of the words in contexts, are useful for the learners both in production activities as well as in the act of retention, as the authors observe:

> in empirical studies of vocabulary learning and dictionary use, learning tends to be operationalised as the extent to which new words, meanings or collocations are remembered by the learners in immediate and/or delayed (unexpected) vocabulary tests following decoding or encoding tasks involving the use of a dictionary.

Another aspect that they highlight is the need for some training as regards dictionaries' consultation: students indeed mostly tend to look for unknown words, while searching for more frequent and common words would increase their knowledge of their multiple uses. Textbooks writers as well now take into account more information taken from corpora in their choice of contents and can rearrange structures, topics and grammar points based on their actual frequency in the language, also going through the different meanings of forms (as for example of verb tenses). Meunier and Gouverneur (2007), in their investigation about phraseology in advanced EFL textbooks, take into consideration five of them (*Cutting edge*, *Initiative*, *Inside out*, *New Cambridge English* and *New headway*) and make some interesting remarks in this regard: first of all, linguistic phenomena like collocations and idioms are made aware to learners already in the back covers of the books, which denotes a greater interest and understanding of their importance; they are taken into account in many exercises too, where learners are required to match the two parts that form them or the whole expressions with their definitions; in the various sections of the books there are also boxes called 'useful language' or 'language toolbox', which can supplement listening or vocabulary tasks by adding other effective utterances in the given contexts. There is therefore a growing commitment to vocabulary (and not only to grammar) and in particular to more phraseology aspects rather than single words in English textbooks. On the other hand, Römer (2004) highlights how English textbooks in Germany still rely on invented sentences, especially in dialogues, which probably never occurred in any speech: the use of real examples taken from speech corpora, which if too elaborate could be modified according to the level of the coursebook, would thus make more sense, since it would provide learners with natural language which would actually be useful for their future communication. As Mindt points out (1997: 46), "without distributional data there can be no informed grading of the functions of a grammatical form in a language course": relying on genuine data should therefore be the usual way in creating grammars and language materials, instead of through a tradition of teaching which often does not reproduce the authentic language use.

Besides the creation of these tools, corpora can also be used directly in the classroom: the language samples which they contain can show how words are actually used in the various contexts, and this can represent an engaging learning procedure for the students and also an easier one for the teachers, who instead of making up examples and thus wasting a

relevant amount of time, can extract representative parts from them according to what they require each time.

The pioneer of this concept is Tim Johns, who coined the term 'data-driven learning' (DDL) as referring to "an approach to the use of corpus data in language teaching and learning" (Johns 2002) in which the learner has a prominent and direct role, since he acts as a researcher himself while recognising facts from the data. As Bernardini notes, in this way learners are guided to discover the language they are learning through "the observation and interpretation of patterns of use" (2004: 16), especially across concordance lines and exercises created from them, and take the more active role of researchers mentioned by Johns. The corpus-driven method can be referred as inductive, since it is directly from the analysis of the data of the corpus that it is possible to make assumptions about the language, and researchers are not influenced by any prior notion (Callies 2015). This approach can bring many benefits such as developing the analytical abilities of the students, who become more autonomous in their own learning and also get to cooperate with their classmates while reflecting and discussing about the data presented, but also as regards the teachers, who coordinate this process, encourage students to be motivated and take part themselves in the research: this can represent an additional incentive for the learners, who don't see teachers only as authorities who give marks and tell rules anymore, but as figures who help them and who can learn from them as well.

Translation can be regarded as an additional activity with educational purposes in the process of learning languages, in that it can also show different patterns of words and expressions in parallel and comparable corpora: their peculiarity is that they contain samples of more than one language, thus enabling learners to make comparisons, find equivalents through the alignment process and discover similarities and differences from their own native language. Chambers (2010: 105) summarises the main advantages noted by researchers by affirming:

> "[…] that bilingual concordance data help learners to avoid the assumption that there is one correct answer, to study words and expressions which are seldom translated literally, to translate idioms and collocations and culturally bound concepts, and to distinguish frequent use from idiosyncratic".

This type of corpora can be productively used also for the so-called reciprocal learning (Hunston 2002, Johns 2002), which consists of pairing a couple of language learners who can benefit from each other (for example, a Spanish student learning English paired with an English student learning Spanish): sentences and their translation in both languages can be extracted from the corpora, so that students can reflect and identify the various ways in which words are expressed and then help and motivate each other in trying to complete others where words are taken out.

Furthermore, the process of selecting extracts from corpora in the classroom can be engaging also because it can lead to unexpected discoveries: as Aston (1997) notices, it is not unusual to work with texts for a particular aim and retrieve also other interesting aspects of the language, which enriches the students' learning and can be another incentive for them too.

1.3.1 Data-driven learning

Exercises based on data from corpora can be of many different typologies. The following research carried out by Johns (1991) focuses for instance on a single verb, in this case *should*: in order to show its various meanings, he sorted the different examples from the corpus in sets, one for each meaning, hence his students were facilitated in the process of seeing and reflecting on the differences of usage and then of denominating the rules based on which the categorisation had been made. In the same essay he focused also on the difference of use between two similar verbs, a problem that students often encounter: he showed that a lot of help can come from concordances of a corpus, because seeing the various contexts in which the word in question is used can generate considerations, in a process which he calls "identify-classify-generalise". The two verbs considered in his study are *convince* and *persuade*: he illustrated how his students quickly noticed that while almost all (10 out of 11) occurrences of *convince* were followed by a that-clause, the ones of *persuade* (14 out of 18) were followed by a to-infinitive, thus making an identification of the two verbs and a classification based on which elements follow them. He then explained to the students that from these results they could generalise that *convince* and a that-clause are used with a change of subject, while *persuade* and a to-infinitive are used with no change of subject, but they had another interpretation in mind:

from the evidence of the examples in the concordances, they generalised that the to-infinitive refers to actions, while the that-clauses to truths, and therefore that "we typically persuade someone to do something" and "convince someone that something is the case". This is a significant example that shows how students can infer or test rules through the use of concordances; in Gavioli's words, they "[…] can examine concordance materials to work out features of language use on their own" (2009: 43), in a way that engages them and makes them more responsible. Gavioli herself conducted a research study with students using concordances from a corpus (2009): in this case, the focus was not on verbs but on genres and they analysed ads and film reviews. In the latter case in particular, while searching for the most common patterns used to express movies' evaluation, they made interesting discoveries about how that is different based on the gender of the performers: while the reviews highlighted the skills and talent of actors, with actresses the judgements were related only to their physical appearance. Students therefore independently interpreted the data and came to their own conclusions, playing an active role in the research and contributing with something useful about the inequality of genders in society that had not been discovered yet.

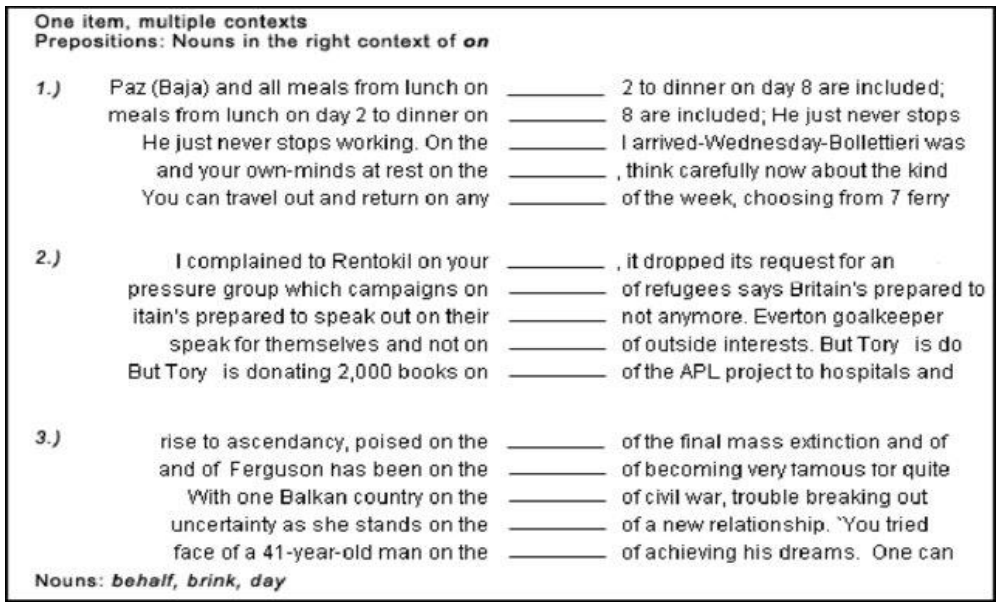
The starting point of students' searching through corpora can also be mistakes they made themselves: an investigation made by Gilmore (2009) shows in fact how corpora can be useful in improving writing skills. The students involved were at their second year of University in Japan and were taking part in a writing course: any errors in the first drafts of their productions were highlighted by the teacher and after a 30-minute introduction on how to use online corpora and a subsequent hour spent working with the BNC and COBUILD corpora in particular, they were asked to write a second draft in the attempt to correct their previous mistakes by looking at authentic sentences. Out of 350 linguistic problems identified in the first drafts, 214 resulted more natural in the second ones, leading to the result that "learners are able to make corrections based on concordance evidence" (Gilmore 2009: 367); furthermore, feedback on the activity showed how the majority of students (95%) was satisfied with the procedure, specifically with its quickness and the possibility of correcting their own texts.

Concordances from historical corpora can lead to a diachronic study of the languages: making a comparison between different periods of time, it is possible to identify changes

in the meaning of words and in their lexis (for example looking through affixes or verb endings), making discoveries about their origins (Knowles, 1997).

Exercises can also involve the area of preposition usage and can be of the 'fill in the missing preposition' type, where students are introduced examples of concordances where the prepositions in the sentences are taken out and need to be put back correctly (giving all the prepositions at the end of the page, in a random order).

A similar exercise can be the one illustrated in Figure 2 called 'one item, multiple tasks' shown by Johns (2002), where, instead of prepositions, lexical words are taken out from the sentences and then listed at the end of the page: what is interesting here is that each word has to be put more than once in consequent sentences, in order to show the range of meanings of the words based on the contexts and to train the learner in formulating hypotheses.



One item, multiple contexts
Prepositions: Nouns in the right context of *on*

1.)
Paz (Baja) and all meals from lunch on _____ 2 to dinner on day 8 are included;
meals from lunch on day 2 to dinner on _____ 8 are included; He just never stops
He just never stops working. On the _____ I arrived-Wednesday-Bollettieri was
and your own-minds at rest on the _____ , think carefully now about the kind
You can travel out and return on any _____ of the week, choosing from 7 ferry

2.)
I complained to Rentokil on your _____ , it dropped its request for an
pressure group which campaigns on _____ of refugees says Britain's prepared to
itain's prepared to speak out on their _____ not anymore. Everton goalkeeper
speak for themselves and not on _____ of outside interests. But Tory is do
But Tory is donating 2,000 books on _____ of the APL project to hospitals and

3.)
rise to ascendancy, poised on the _____ of the final mass extinction and of
and of Ferguson has been on the _____ of becoming very famous for quite
With one Balkan country on the _____ of civil war, trouble breaking out
uncertainty as she stands on the _____ of a new relationship. `You tried
face of a 41-year-old man on the _____ of achieving his dreams. One can

Nouns: ***behalf, brink, day***

Figure 2: One item, multiple tasks (Johns 2002)

Vocabulary acquisition is thus another area in which corpora can be extremely useful: Donesch-Jezo (2013) explains how the advent of computerised database in the new century was essential in a shift from only grammar to vocabulary too in English teaching, since before the assumption shared by researchers was that grammatical items were the ones to learn at school, while lexical ones could be acquired later according to each student's needs. She conducted a study with intermediate-level learners of medical

science using research articles written by native speakers in the field of medicine; they were made aware of what a corpus is and how a concordancer works and before the analysis of the texts they created a wordlist that showed the frequency of occurrence of all the words contained, in order to decide the ones to focus on first. Starting from these, they looked at how they were presented in the concordances, the most common words that preceded or followed them (in particular the correct prepositions that occurred before and after) and the others that together formed collocations and after having taken notes, they made generalisations about their findings. At the end of the study, a test could prove how the students improved their vocabulary, and a feedback from them showed that the procedure was interesting and useful for understanding data from their field of study.

Quoting Cobb (1997: 314), "in learning a second language, there is simply not the time, as there is in a first language, for rich, natural, multi-contextual lexical acquisition to take place": corpora offer therefore one solution for this problem, as they present the learner with vocabulary input in its natural context and in a quick way, favoring the noticing and the mental process of memorisation of words.

Tribble and Jones (1997) also mention the use of concordances in the study of homonyms and synonyms, where an example of exercise can be the one of identifying and separating the different meanings or functions, which can be done individually or even as a group activity. In this way, as they put it (Tribble & Jones 1997: 7), "the study of grammar (or vocabulary, or discourse, or style) takes on the character of research, rather than spoon-feeding or rote learning". Hunston (2002) points out that tasks of this type can be integrated in the classroom starting from a word or a phrase met during the lesson, for example while reading a text: concordance lines of the word in question can be selected and shown to the students, who afterwards can complete exercises with filling gaps or even predicting what other words could have the same pattern.

A relevant aspect that has to be considered, underlined by Gavioli (1997), is that giving students a concordance software together with data is not enough to make them researchers: the difficulties that they may encounter do not only regard the unfamiliarity with this type of technology, but also the analysis and identification of patterns, for which a specific linguistic knowledge is required. Therefore students need to be introduced and lead gradually in this process, and made aware of the fact that data are the starting point and not the result, since not only they often do not provide immediate answers, but also

they can cause more doubts and questions instead: this is the main aspect in which the use of concordance lines differs from the one of grammars or dictionaries: only the former need to be read at various levels and interpreted through hypotheses, while the latter, offering a description, can be used in a subsequent moment to confirm possible intuitions. Only when the learners acquire more technical skills and become more autonomous, then can they also engage in topics of their choice that they find particularly interesting: this last concept has been highlighted by Boulton (2019a), who describes a course that is part of a master's degree in a teaching program in France.

The task that the students were required to do was to choose a topic and the relative questions of their interest, then compile a corpus for analysis and lastly write a paper about their findings. Some of the titles collected revealed how the areas of topics could vary from literature ("The importance of invented words in the Harry Potter books") to cultural studies ("Societal notions in the same-sex marriage debate in the United States of America") etc, showing the variety of aims for which corpora can be used; a personal feedback section compiled at the end of the course revealed also a general appreciation of the assignment from the learners, who reported that they had gotten involved and motivated after seeing their first results and that they would enjoy further research with the same methods in the future.

1.3.2 Data-driven learning across levels

It needs to be noted that the authors mentioned above in the exercises' section specify that the learners involved in the research were advanced; considering the fact that in modern societies the majority of language learners are not university students (Pérez-Paredes 2010), more attention needs to be brought to them, adapting the methods in the field of research to the ones in language education. With the mediation of the teacher, who can extract examples and think of appropriate tasks, it is possible also for these learners to take advantage of corpora. This means that pedagogy needs to be involved regarding the transfer of corpora from research to school: in fact, much literature has illustrated their positive effects on learners, but it is essential to point out that they need to be adapted in the learning environment in order to do so: students are not at the same linguistic level as the linguists who work with corpora, and for this reason it is the teacher's job and responsibility to determine whether and what findings could be of use to them. Sinclair

(1987 in Aston 1995: 259) summarizes this in describing a corpus as "a database for teachers' reference, a repository of facts about [English] on which new syllabi and materials can be based", and therefore not as "a source of data for use by students themselves" (Aston 1995: 259). Consequently, teachers need to be introduced to corpora and their usage since the early stages of their profession, as some authors point out (see Boulton 2010; Braun 2010) already in teacher-training courses, also to acquire knowledge in information and communication technology. One way to adapt corpora to the learning environment is through tagging, which is defined as an addition of a code in a corpus, which can be much easier if computer-assisted: teachers can take extracts or texts and add annotations playing a mediation role, with a pedagogic intention of explaining a concept linguistically, thus adding value to the corpus, or as a basis for creating exercises for their students, "as a way to bring the corpus closer to the learner", as Pérez-Paredes puts it (2010: 65); a valid example of this procedure is the project realised by Sealey and Thompson (2004) with primary school learners from the age of 8 to 10. These children were introduced to the concepts of corpus and concordance through the use of tagging, as every word was coded with its relative part of speech, and in addition each one was highlighted with a different color, in order to make the process more involving in consideration of their age; the sentences and words taken from the corpora were simple structured and regarded topics that the students already encountered in class. Recordings of these sessions of study could prove that this approach was appreciated and not considered difficult by the students: when they were asked to describe the work that they had been doing, they easily referred to the word 'concordance' and the role of colors in showing different meanings. Furthermore, based on the different colors, they also valued some words as more exciting than others (that is, open-system words like nouns and verbs, opposed to closed-system words like articles and pronouns), which revealed how this method was useful in this type of categorization too.

The author also mentions the so-called 'feasibility scenario' (Pérez-Paredes 2010), which, in order to work, would need besides the involvement of teachers also the commitment of corpus linguists, who would try to develop themselves resources adaptable to more educational contexts: this would considerably help teachers in their job, since they have to follow school textbooks and often have no time left for extra activities. As a matter of fact, what emerged from a questionnaire addressed to English teachers at secondary

schools devised by Römer (2009), regarding any difficulties in their work situation, is that they often have to create materials themselves wasting much time, and that in marking students' essays or exams they often cannot find adequate answers in dictionaries about the acceptability of word combinations. Moreover, among other resources used more frequently, they refer to the possibility to ask directly a native speaker, which is clearly not available or immediate for everyone, and Internet consultation, which often poses the problem of the authenticity and reliability of its copious sources. Since teachers' complaints about textbooks in the comments section of the questionnaire regard above all the lack of language variation, up-to-date material and adequate exercises, an integration made from corpus researchers of data-driven learning activities into these resources represents an ideal solution to all these results and would be convenient for everyone in the learning environment.

The collaboration between these two figures would therefore be beneficial: as Boulton (2019b) claims, the lack of communication between them leads to the majority of studies regarding data-driven learning being taken in the university environment and not being known in the educational system, also because of the general perception that the linguistic level of younger learners is not good enough for them to take part in that type of activities. Actually it should also be taken into consideration that the youngest generations, having grown up with the Internet, are already familiar with the act of searching and interpreting data from a large source; this can represent an obstacle too, as the study presented by Geist and Hahn (2012) shows: the two authors worked with intermediate learners from the age of 14 to 16 and noticed how their familiarity with Internet can be disadvantageous since having a computer at their disposal distracts them from the linguistic task; furthermore, knowing that their work would not be evaluated was another factor that decreased their motivation. More research should thus be conducted on high school learners, to see if these are general common problems typical of their age.

As for University students, a meta-analysis of corpus use in language learning (Boulton & Cobb 2017), which aimed at showing an overview of the effectiveness of the tools and techniques of corpus linguistics for second language learning, at the end could prove that data-driven learning activities are appropriate for undergraduates as well as graduates and for intermediate level as well as advanced.

Beginning-level students were taken into consideration too in an investigation conducted by Chujo, Utiyama and Miura (2006): the authors acknowledged how concordances can be overwhelming for inexperienced learners because of their length and complexity and made use of a Japanese-English parallel corpus in order to facilitate their learning process, in this case focused on vocabulary. Furthermore, as a multilingual concordancer they opted for *ParaConc*, which includes an ulterior element called 'hot words' that displays the possible translations for any word based on frequency and therefore makes the process of choosing the right equivalent easier.

The activities involved a small quantity of data, which were not ambiguous but rather easy and precise, and consisted of the examination of Japanese equivalents of an English word, the examination of English equivalents of a Japanese word, the collection of frequent citations of English words, the translation of Japanese phrases into English, the examination of collocation patterns and finally the generalisations of recurrences, for a total of five weeks. The students worked in groups and were thus introduced to the many existing relationships between the two languages and were guided in the discovery of the most typical English expressions and their meaning, also through the integration of the CALL program. The results showed that they could manage to use the technological tools, with only 16% of them who reported to have had difficulties; a final questionnaire in addition showed how they found the tasks useful and interesting for vocabulary learning and that they would have liked to use it in the future as an addition to vocabularies in the process of searching for words and examples. Parallel corpora proved therefore to be excellent tools for beginners too, who need by any means to be guided adequately, for instance by providing an example at the beginning of each task, as the authors suggest.

In the end, when using corpora in the classroom, "the aim does not necessarily have to be to make proficient corpus users of all students" (Granath 2009: 49), but to help them in their learning process providing authentic examples, which can be selected according to their linguistic level; it is therefore possible in every educational environment.

1.3.3 Hands on versus hands off data-driven learning

Another aspect that characterizes the exercises cited above is their use in the classroom in printed form: both students and teachers have in fact claimed that the use of software can represent a barrier, as regards the availability and functionality of computers and

computer rooms, the difficulty in understanding how they work (as they are mostly designed for the university environment) and the gap mentioned above between the methods used in research and their application in language education.

However, researchers have wondered if it is possible to gain the same achievements with paper-based materials and with electronic ones, and if the two approaches can be considered valid alternatives or if there are significant differences.

Chujo and Oghigian (2012) illustrate how both present advantages for the learners: paper-based materials do not require an ability in working with technology and can therefore save time and be easier to use, while at the same time only computers provide a wide amount of data and stimulate skills and strategies which can be useful for numerous other types of tasks; they conducted therefore two studies with students who used both approaches alternatively, in order to have a feedback from them regarding a potential preference between the two. The participants were Japanese learners at the beginner level and the activities regarded the identification of particular word classes or phrases with the aid of a parallel Japanese corpus; a test which they took before and after the studies established how the activities were effective, however with no significant gap between the two methods. The feedback from a questionnaire did not provide a clear propensity for one method either, as students reported that paper-based resources were easier and quicker, while at the same time computer-based ones were more engaging and better for memorizing new data: a blending of the two is thus what the authors suggest, meaning their use in different steps of the students' learning (printed materials first and subsequently the use of computers).

Boulton (2012) conducted a similar study too, but with lower-intermediate level learners (second year of study at an architectural college), in order to see if the learning results varied from students who worked with printed exercises (hands-off approach) and students who worked with computers (hands-on approach) and also if the students themselves showed a preference. They were divided into two groups and they all experienced both methods on alternate weeks for a long period, which made possible their familiarity with both procedures and a comparison between them in the end; the activities regarded the interpretation of data from concordances, gap-fill exercises and further research that could be conducted starting from them. The results showed that both methods proved to be effective for the learners and only a slight advantage for paper-

based activities; the questionnaires created in order to obtain a feedback from them were made of sentences related to the use of corpora with which the students could agree or disagree based on a 5-point Likert scale: these showed a general receptivity for both approaches, with a slight preference for computer-based activities. Therefore, as there was no significant difference between the two groups' scores and preferences, an interesting compromise would be to not rule out any of the two, but to try to use them together in order to exploit both their features and to facilitate the students in the process of learning: including exercises taken from corpora into printed materials would save precious time, as working with technology and the problems that may derive from it always require much of it; considering also the prior training which could be required, it would therefore be more appropriate for learners of all levels too.

This concept is shared by Frankenberg-Garcia, who affirms that the two methods "are not mutually exclusive, and there is a time and place for both" (2012: 49): in fact, instead of using them in different moments of the learning process, she examines how they can be used for different aims, being extra activities for the whole class for printed materials and a detailed analysis for single students, who may have individual doubts, for online corpus consultation. As examples, she mentions how printed concordances and collocations can be given to the class before practicing conversation or a writing task, in order to widen their vocabulary with authentic language that is not present in their regular coursebooks; at the same time, the same students can look up words or sentences in online corpora during and after their production to check the correctness of their texts, in a way that results more accurate than through dictionaries. Corpora are thus a potential supplement to all kinds of classroom activities, as well as to textbooks, to the advantage of both learners and teachers, by stimulating and providing new input to learners from beginner to advanced level and for teachers in the preparation of appropriate materials and in the correction of their students' work.

In conclusion, the use of DDL activities should be more publicized, as they can be useful integration and extension to the already existing techniques used for teaching.

Much evidence has shown how "corpus use needn't be complicated or separate from everyday teaching, and can be realistically integrated with set course materials and well-known classroom activities", as Frankenberg-Garcia states (2012: 39). The integration of samples from corpora in published materials like coursebooks would make them more

relevant and would make them reach a wider audience, as Boulton (2010) further points out, besides the advantage of ensuring a variety of learning methods and the one of removing those technical aspects which might seem too demotivating and long to learn.

CHAPTER 2

2. LEARNER CORPORA: A MORE DETAILED ANALYSIS

After a general introduction to corpora in the previous chapter, I now focus only on learner corpora, providing a more detailed description and results from studies based on them. These studies refer for the most part to English texts produced by Italian and German learners of English, the two categories which I will take into account in my research in the final chapter.

2.1 An overview of learner corpora

The main characteristic of learner corpora is that they are made of collections of data produced by foreign or second language learners, and therefore they can be a useful aid in the process of learning languages for different reasons as compared to corpora produced by natives. In fact, native corpora are helpful as they present the authentic language, but on the other hand they do not indicate which structures are the most difficult to learn for students; through learner corpora instead it is possible to investigate how learners use the target language, and what they get right is equally important as what they get wrong, since both can help to understand how that language is learned and what can be improved. The data collected are formed of naturally-occurring texts as opposed to experimental ones (such as multiple choice tasks), as they are more pertinent for the study of what learners produce spontaneously and as they allow an investigation of more aspects at once: these may be, for example, essays (in written corpora) or oral interviews (in spoken corpora) produced by students at an intermediate level.
The criteria considered for the collection of data allow a more precise categorisation of studies, and include among the others (Meunier 2016: 377):

> information on the home country, the native language, the various language(s) spoken at home, the knowledge of foreign language(s) and their respective proficiency levels (as self-assessed or as officially assessed by educational organisations or internationally recognised proficiency

tests), the language(s) of instruction in the learner's educational background, the amount of target language instruction (e.g. number of years spent studying the target language), the type and amount of input received in class or during extracurricular activities, the time spent in a country where the target language is the official language, and the potential use of reference tools (dictionaries or other) to prepare the text.

Initially most learner corpora were corpora of L2 English, however there is an increasing number of learner corpora of other languages too: examples are the written learner corpus of French *French Interlanguage Database* (*FRIDA*) mentioned by Nesselhauf (2004), the written learner corpus of German *ALeSKo* mentioned by Breckle and Zinsmeister (2013) and the spoken learner corpus of Italian *Lexicon of Spoken Italian by Foreigners* mentioned by Gallina (2013).[1]

It is vital to be cautious when inferring from results of this typology of corpora: as Granger (2004: 125) affirms, "Learner language is highly variable. It is influenced by a wide variety of linguistic, situational and psycholinguistic factors, and failure to control these factors greatly limits the variability of findings in learner language research". Examples of these variables that may influence different outputs can be the type of education and the cultural context, the learners' age and gender, as well as their mother tongues.

Much research with learners' data concern in fact the level of influence of their native language and the areas of overuse, underuse and errors typical of native speakers of a language compared to native speakers of another one (different interlanguages), as well as compared to native speakers of the target language studied (native language vs. interlanguage). These two types of comparison form the so-called Contrastive Interlanguage Analysis (Granger 2009): the first allows a better study of any peculiar features shared by a certain interlanguage that are then reproduced in the learner output, while the second highlights the aspects in which native and learner language are similar or differ, and thus requires a native corpus of the same genre in order to compare the two.

The other approach to study learner corpora is Computer-aided error analysis, which can consist either of selecting a linguistic item considered particularly difficult and then retrieving all its instances in the corpus through a software tool, or directly of tagging all

[1] For a complete list, visit the website https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html

the errors, or the ones of a determinate category, through a system of error tags. The analysis of errors does not have to be regarded as negative, as they are a source of information that can help learners to improve (Granger 2002).

As Granger (1998) explains, the first learner corpora were very small (around 2000 words) and were used only with the aim of extracting the main errors; then, with the advent of computers, more data could be analyzed and compared and more phenomena studied. The author herself is also the founder of the most established computer learner corpus, the International Corpus of Learner English (ICLE): this is made up of written essays, which are about 500 words long, produced by young adult advanced learners of English as a foreign language of 16 different native tongues (Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana and Turkish). The high number and length of essays from different nationalities can thus allow many qualitative and quantitative studies about many aspects of the English language produced by learners, often with the use of the LOCNESS Corpus too as a reference for native-speaker data.

Computer learner corpus research, the linguistic methodology which consists in collecting learner data electronically in order to study them, is particularly advantageous in the two fields of Foreign Language Teaching (FLT) and Second Language Acquisition (SLA). The difference between the two is that while the first "favours the study of contexts of guided learning through the observation of actual individual or classroom learning situations […]" (Prat Zagrebelsky 2004: 39), the second aims to "understand how people acquire another language spontaneously and discover what is first and most easily learnable and why" (ibid: 39); a collaboration between SLA specialists and corpus linguists would therefore be beneficial, as it would provide more quantitative data to test SLA hypotheses.

In addition, studies can either focus on the differences and similarities between many non native learners or between different periods of time among a more restricted number of learners: depending on these criteria, they can be classified as cross-sectional, i.e. "they contain data gathered from different categories of learners at a single point in time", or longitudinal, i.e. the "data gathered from the same learners are collected over time" (Granger 2004: 131).

With learner production it is essential to get a wide amount of data and people, in order to be able to define a particular phenomenon as characteristic not only of a few learners, but typical of a whole category; considering also how the process of manual transcription is even more time-consuming with learner data compared to native ones, as it is particularly relevant to ensure that every potential error is reproduced in the same way as in the original text, it is fair to say that the possibility to computerise these data has benefited this type of studies (Granger 2004).

Other functions that are facilitated by automatic tools are the count, sort and compare ones, as well as annotation: although it is possible to do research on raw learner corpora, this "practice of adding interpretative, linguistic information to an electronic corpus of spoken and/or written data" (Garside et al. in Granger 2004: 128) is particularly relevant in the learner type, since the identification of specific features can result in a better and more systematic work. For example, part-of-speech (POS) tagging, which allows to assign to each word in a corpus a tag indicating its word class, facilitates research on specific categories that were previously difficult to explore.

The most relevant form of annotation however is error annotation, since it can provide a picture of how language acquisition works by observing which type of errors learners make and how they make progress at different ages and stages; since errors can be of many types, as well as the error tagging systems, in order to achieve this and avoid subjectivity it is essential to consider some tagging guidelines (Diaz-Negrillo & Thompson 2013, Tono 2003).

Reznicek, Lüdeling and Hirschmann (2013), in describing the German learner corpus Falko, refer also to the need for multiple target hypotheses, which are the correct forms of any potential errors; they also claim that a multi-level stand-off architecture in annotation, in which each layer is independent from the others and stored in a separate file, is the best and most flexible one, as it allows a clearer interpretation of data.

The authors make a pertinent example to explain this concept: in the sentence produced by a learner *one can still remember Billie Jean King, woh was tenniswoman in the 80 and who fought for one free homosexuality*, there are many errors which involve different areas and thus more than one target hypothesis is required. In this case, some regard orthography (*woh* → who, *80* → 80s), another one grammar (*tenniswoman* → a tenniswoman), another one lexis (*tenniswoman* → tennis player) and another one

semantics (*one free homosexuality* → a free homosexuality). It is not a rare phenomenon with learner data to need more than one target hypothesis in a single sentence, and for this reason placing the correction right next to the error in the original file makes the analysis more difficult and confusing.

2.2 Pedagogical applications of learner corpora

Learner corpora too can be used in the classroom as a tool to create exercises and to enhance the study of languages. The approach of data-driven learning can thus be applied, with some differences as compared to native corpora: learner data can be useful as they present negative evidence, which can be a positive aid for language acquisition since learners, looking for and studying mistakes, acquire a more positive attitude towards them, which become "no longer merely a feature that has to be corrected, but also a feature that can be discovered" (Nesselhauf, 2004: 140).

Moreover, this increases also the autonomy and motivation of learners, who develop their ability to detect mistakes and analyse texts. Nesselhauf (2004) also mentions how this procedure is particularly useful for points that are fossilised in the learners, which means aspects that have been repeated more than once but that they do not seem to get right, since "instead of being told once again that what they are doing is wrong, learners have the opportunity to get something right, namely to identify and explain the mistake in question" (Nesselhauf, 2004: 140).

What is essential when dealing with data-driven learning with learner corpora is that positive evidence from native corpora has to be shown to students too, otherwise they may face the risk of memorising the mistakes they identify: in order to avoid this, exposing them to further exercises can help them consolidate all the correct forms.

Diaz-Negrillo and Thompson (2013) refer to *delayed pedagogical use* of a learner corpus, which consists in the compilation of a corpus from data produced by the learners that can be later used in the design of pedagogic materials like dictionaries and coursebooks, as well as by the teachers. As De Cock and Paquot (2010) remark, the use of learner corpora in the compilation of monolingual learners' dictionaries is relatively recent, with the first one published in 1993 (the 'Longman Language Activator'). The information assembled from them involves on one hand the misuse of target language words and phrases, and on

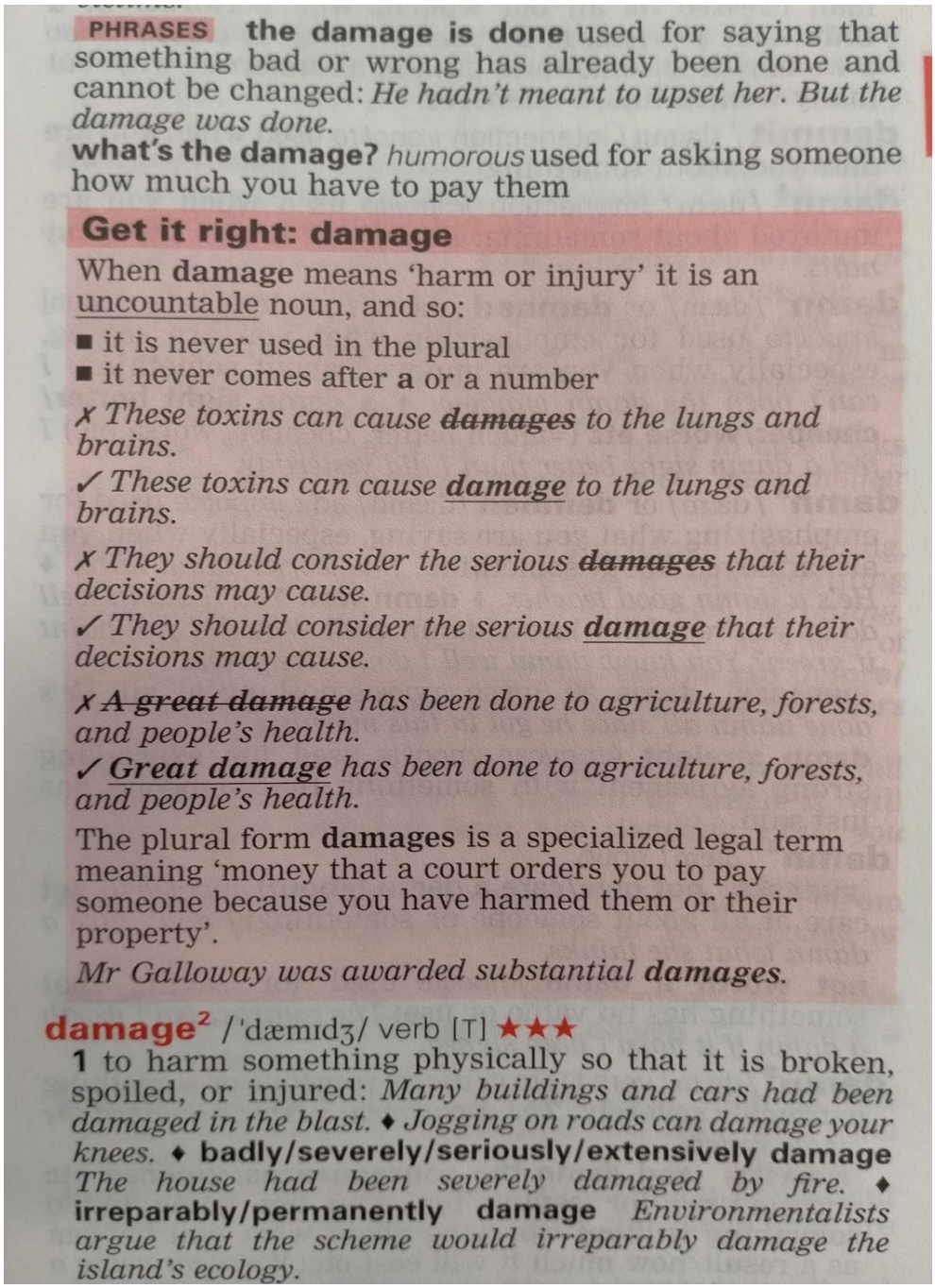
the other hand their overuse or underuse (De Cock & Granger, 2004). In particular, both paper and electronic dictionaries contain language notes and 'help boxes' based on learner data, added to help students in the identification of the most frequent mistakes along with their respective correct usages.

Gillard and Gadsby (1998), in analysing a later monolingual learners' dictionary they compiled called 'Longman Essential Activator', recognise four areas of most typical errors: collocational (*she said a joke and I laughed instead of she told a joke and I laughed),* grammatical/syntactical *(I am liking it a lot instead of I like it a lot),* spelling *(we hear it everyday in the news instead of we hear it every day in the news)* and word order *(I am tall 180 centimetres instead of I am 180 centimetres tall).*

This shows how, as well as for lexis, sections for grammar are also present in these dictionaries, as some aspects might have been forgotten or found confusing by students.

Figure 3 taken from 'The Macmillan English Dictionary for Advanced Learners' analysed by De Cock and Paquot (2004) shows how these 'boxes' are structured, firstly with a description of the word/aspect considered and secondly with some examples,

which aim to provide both the incorrect and the correct version to learners: this "helps them to see what the error looks like and it lets them compare their own sentence with the two examples to check if their own sentence is right or wrong" (Gillard and Gadsby 1998: 168). Considering that the data used are produced by already advanced learners, this proves to be truly helpful, as these mistakes may be very difficult to eradicate.



PHRASES **the damage is done** used for saying that something bad or wrong has already been done and cannot be changed: *He hadn't meant to upset her. But the damage was done.*

what's the damage? *humorous* used for asking someone how much you have to pay them

Get it right: damage

When **damage** means 'harm or injury' it is an uncountable noun, and so:

- it is never used in the plural
- it never comes after **a** or a number

✗ *These toxins can cause* ~~***damages***~~ *to the lungs and brains.*

✓ *These toxins can cause* ***damage*** *to the lungs and brains.*

✗ *They should consider the serious* ~~***damages***~~ *that their decisions may cause.*

✓ *They should consider the serious* ***damage*** *that their decisions may cause.*

✗ ~~***A great damage***~~ *has been done to agriculture, forests, and people's health.*

✓ ***Great damage*** *has been done to agriculture, forests, and people's health.*

The plural form **damages** is a specialized legal term meaning 'money that a court orders you to pay someone because you have harmed them or their property'.

Mr Galloway was awarded substantial ***damages****.*

damage[2] /'dæmɪdʒ/ verb [T] ★★★

1 to harm something physically so that it is broken, spoiled, or injured: *Many buildings and cars had been damaged in the blast.* ♦ *Jogging on roads can damage your knees.* ♦ **badly/severely/seriously/extensively damage** *The house had been severely damaged by fire.* ♦ **irreparably/permanently damage** *Environmentalists argue that the scheme would irreparably damage the island's ecology.*

Figure 3: 'Get it right' box for *damage* from 'The Macmillan English Dictionary for Advanced Learners'.

Concerning coursebooks, Meunier (2016) makes reference to two grammar books where information from learner data has been integrated: 'English Grammar Today' presents some sections enriched with information about the typical errors made by learners (e.g. in the unit dedicated to the present perfect tense, examples are shown where learners tend

to incorrectly use the past simple), gathered from the Cambridge Learner Corpus (CLC); the 'Grammar and Beyond' coursebook, also based on research on the CLC, identifies instead the differences between the grammar of written and spoken English, focusing on the most commonly used structures.

As for teachers, starting from their students, they can observe which aspects of the language they teach are understood and which need more practice, and then decide to create materials from the corpus' data. Activities may involve error correction, error identification or a comparison between concordance lines from a learner and a native speaker corpus, which can show not only possible mistakes, but also an overuse or underuse of any particular expressions; as Nesselhauf (2004) explains, if the learner data are presented after the native speaker ones they generate a sort of surprise, as students can observe how very often a feature is not as obvious as it seems, and in this way they reinforce their understanding.

An example of this type of task is mentioned by Granger and Tribble (1998): after noticing from learner data that the adjective *important* is overused, teachers can decide to enhance their students' vocabulary by creating activities which involve the use of other underused synonyms like *significant*, *major*, *vital*, etc of which they may not be aware; the same procedure can also be applied with different meanings of single words which may be generally less known.

As well as what to teach or improve, learner corpora can help teachers also in giving indications on how to teach or improve particular points: an analysis carried out by Granger (1999) for example draws the conclusion that tenses in English should be taught at discourse-level rather than at sentence-level, as since "students have been taught tenses at sentence level and continue to 'function' at that level even when they have reached a more advanced level of proficiency" (Granger 2009: 197), when they are required to produce a discourse it results in a combination of correct, but incoherent sentences which do not generate an appropriate result.

Furthermore, teachers can benefit from these data also in the creation of profiles of students' performances, a process that can help them with their grading. Granger (2004: 137) also mentions the so-called CALL (Computer Assisted Language Learning) programs, which involve pedagogical activities meant to help both students and teachers through the use of computer networks: for instance, students can write their texts in an

interface and then send them to teachers in order to be corrected, getting therefore an immediate feedback in the form of lists of errors they produce and at the same time providing their professors with additional data. Another program mentioned by Milton (1998) displays instead authentic error types which users must correct, aided by an online grammar linked to every error, while Heift's 'E-Tutor' (Meunier, 2016) contains an artificial intelligence component which provides an evaluation of learners by considering their past and current performances in relation to specific activities.

Opposed to the delayed one, Diaz-Negrillo and Thompson mention the *immediate pedagogical use* of a learner corpus too, which consists in providing the learners with the data they produce themselves: in this way, "learners are at the same time producers and users of the corpus data" (Granger 2009: 20) and since they start from their specific errors, they have the possibility to reflect on how they use the language and are more motivated to overcome the actual difficulties that they encounter.

Teachers can also build a corpus out of the texts written by their own students as a starting point for further activities, with the only difference that, if presented to the whole class, the texts should be anonymised to avoid any embarrassment or derision in the classroom environment.

2.3 Learner corpora based research

Many published learner corpus studies, as Nesselhauf (2004) notes, are based on the various ICLE sub-corpora: for this reason, multiple aspects of each language represented have been analysed, starting from advanced learners' writings.

These studies concern thus aspects peculiar to the learner language: for example, Granger & Rayson (1998) in an analysis of French learners' production of the English language, demonstrated how they tend to overuse more informal words (like the nouns *people, thing*, *problem* or the adverbs *also*, *only*, *so, now*) and on the other hand how they tend to underuse more formal ones (like the nouns *issue*, *argument*, *claim* or adverbs ending in *-ly* like *greatly*, *importantly*, *currently*). In addition, generally "the non – native-speaker writers overused three categories significantly: determiners, pronouns and adverbs, and also significantly underused three: conjunctions, prepositions and nouns" (Granger & Rayson 1998: 123). Another study carried out by Altenberg & Tapper (1998)

with Swedish learners showed many similarities with the previous one, which seems to prove how these tendencies are not influenced by the respective native languages.
Another lexical aspect that has been analysed is verbs, especially those used in academic discourse (Granger & Paquot 2009). The authors take into consideration a previous classification of them into five categories, namely activity verbs (e.g. *give*), reporting verbs (e.g. *suggest*), mental/emotive verbs (e.g. *know*), linking verbs (e.g. *prove*) and logico-semantic relationship verbs (e.g. *follow*), along with the most frequent elements with whom they are correlated, mainly nouns and adverbs; in fact they explain (Granger & Paquot 2009: 196) that

> lexico-grammatical restrictions of EAP (English for Academic Purposes) verbs are often disregarded in EAP textbooks, which tend to present verbs separately from nouns and adverbs when in fact, as demonstrated by several recent learner corpus-based studies, it is their interaction that causes difficulty for learners.

The results, obtained from data from the ICLE sub-corpora and ACAD, a collection of native expert writing, refer to both verb lemmas and verb forms, in order to detect not only what verbs the learners know, but also the phraseological patterns of use they employ. Concerning the former, out of the top 100 in both corpora, only 35% are shared by both of them, with differences in ranking and frequency as well (with significantly more overused than underused verbs by learners, in particular activity and mental ones). As for the latter, it was found a wider area of overlap but also a distortion regarding some differences between the two corpora: the authors mention as an example the verb *conclude*, which at the lemma level displays no differences in frequency, but which shows instead an overuse of its infinitive form (*conclude*) along with an underuse of its third person singular of the present simple tense (*concludes*) plus its past simple form (*concluded*) in the learner production. This is mainly due to their overuse of the connector *to conclude*, while native speakers tend to use equally this form and the subject-verb one (e.g. *we may conclude that…*). Therefore, from these results it is possible to affirm that learners should be made aware not only of more alternatives of the most frequent verbs, as these ones result overused, but also to the many forms in which these verbs can be used in academic writing tasks.

Granger (1997) studied also grammatical areas, such as that of non-finite clauses, namely clauses whose verb element is non-finite, in particular participle ones (e.g. *Returning to my village after thirty years, I met an old schoolteacher*) comparing data from the sub-corpora of ICLE and from LOCNESS (Louvain Corpus of Native English Essays) for the native-speakers counterpart. After dividing participle clauses into nominal, adverbial and postmodifying ones, she analysed each category and found out that their frequency in learner corpora is statistically significantly lower as compared to native speakers' data, precisely half as many; despite this, an interesting fact is that the two corpora present the same distribution of -ing (two thirds) and -ed (one third) participles. As key factors to explain this phenomenon, the author refers to EFL grammars, which do not focus particularly on this grammar aspect, and to the fact that these clauses are often not present in the other languages, making thus the process of learning them very difficult: in fact, the learner corpus presents a percentage of misuse of these clauses too.

However, in recent years, many other sources are being used to study different aspects of learners of English in addition to the ICLE: examples are the Longitudinal Database of Learner English (LONGDALE) (see Castello's study in 2.3.1) described by Meunier (2015) or the more specialised Academic Learner English (CALE) (see Wiemeyer's study in 2.3.2) mentioned by Callies & Zaytseva (2013) and Advanced Learner Corpus of Argumentative Student (ALCASE) mentioned by Cresswell (2015).

2.3.1 Italian learners of English research

Prat Zagrebelsky (2004) reports five studies on different linguistic areas based on the Italian sub-corpus of the ICLE (ICLE-IT) compared to the native corpus LOCNESS. The first one regards the use of contractions in the writing of academic essays and shows how Italian learners of English overuse contracted verb forms and with a higher frequency as compared to natives. The author takes into consideration both the frequency of contractions and the so-called "ratio of contraction", which is "the proportion of contracted forms versus full forms that are potentially contractible" (Prat Zagrebelsky 2004: 104): in particular, salient cases are found to be the ratio of contraction of the negative form of do (*don't* instead of *do not*), as it is used in all registers, and the one of the first singular person of the verb to be (*I'm*, *I'm not* instead of *I am*, *I am not*), which highlights how learners tend to express more personal opinions than natives, who

instead prefer a more detached tone in this kind of essays.

The contractions in the ICLE-IT are further divided into four groups: forms that are wrongly contracted (e.g. *needn't*), forms with low frequency (e.g. *aren't*), forms with high frequency but low ratio of contraction (e.g. *it's*) and forms with both a high frequency and a high ratio of contraction (e.g. *can't*).

This usage of contractions typical of spoken and informal language on the other hand is not considered appropriate in formal writings and is explained by the author as caused by the adoption of a more participatory style close to the spoken one in newspapers (already mentioned in previous studies of the English press which date back to the 1960s until the 1990s), as well as the impact of the media, which may not make learners aware of the other existing stylistic alternatives more suitable for the academic register.

The second study concerns lexical items, in particular five words belonging to the semantic field of work (*career*, *employment*, *job*, *occupation*, and *work*): although their number of occurrences is the same in both corpora, learners tend to make more errors assuming wrongly that they can be used as synonyms. As for *occupation*, the author in fact reports an example where in the non-native corpus (NNC) it is misused in the collocation *first occupation*, which was probably a literal translation of the Italian *primo impiego,* which could have been better rendered using *job* instead; in addition, in the native corpus (NC) the term refers also to a general activity, not necessarily a job. Differently, the term *employment* tends to refer only to the field of work in the NC, while in the NNC is also used in the sense of "use" (e.g. *employment of unusual words…*): this is not considered a proper error, but it sounds unusual and too formal. As for *career*, an interesting fact obtained from the corpora is that in the NC it tends to have a positive connotation as a source of gratification, while in the NNC is connoted negatively as something that takes up people's energy and time: this shows how cultural factors too are transferred when writing in a foreign language. Moreover, in the native usage it is rather common to use other nouns as modifiers (e.g. *career choice*), while learners often make mistakes trying to produce collocations because of the interference of their nativetongue (e.g. *make a career* as a literal translation of *fare carriera*).

Finally, a wider variety of expressions of *job* and *work* can be seen in the NC, like *to be out of a job*, *work ethic*, etc while in the NNC they are sometimes misused as synonyms (e.g. *heavy jobs, looking for a work*). What emerges is therefore a great difficulty for

learners to understand the differences between these terms, since in their mother tongue they can be used interchangeably: more work should be done with the use of concordances and dictionaries in order to focus on each meaning and distinguish their various usages.

The third study investigates the difficulty of cleft sentences for learners: the results show not only that students make few errors in their structure, but that they also overuse some of them in comparison with natives. The two types of cleft sentences analysed are it-clefts (e.g. *it was George who found the right answer*) and wh-clefts (e.g. *what he's done is spoil the whole thing*): while the first type is used similarly in the two corpora, with a distribution between noun, prepositional, adverbial and subordinate phrases, the second one is more frequent in the learners' corpora. It is also interesting to note how in the native corpus wh-clefts tend to introduce new information, while in the learner one they tend to set apart one element from others of the clause; the only significant errors detected regard the omission of the preposition *by* before the *-ing* form (e.g. *it is not studying that a young person can learn how to listen ...* instead of *it is not by studying that a young person can learn how to listen ...*) and the addition of the pronoun *it* in wh-clefts (e.g. *what is unacceptable it is the reaction of the people ...* instead of *what is unacceptable is the reaction of the people ...*). A reason for this abundance of cleft sentences may be found in the fact that learners are over exposed to this type of structure in their studies of the language and often do not know other devices to assign prominence to specific elements in a clause, like extraposition or thematic fronting, as the author suggests.

A longitudinal study about it-clefts only has been conducted by Castello (2017), who used as a reference the Italian component of the LONGDALE and specifically essays written by undergraduate language students in their first three years of university. The author came to the conclusion that Italian learners overuse this type of structure too if compared to natives: in particular, in their first year they tend to use more the pattern "it (mod) V (adv) adj", as in "*for a language learner it is almost impossible to master a second language*" (Castello 2017: 163) while in their second and third year they prefer the pattern "it (mod) be V-ed", as in "*as just mentioned above, it can be claimed that going abroad is a successful way to know and get in touch with foreign cultures*" (ibid. 165) which is more characteristic of advanced academic writing. Moreover, more types of mistakes could be found due to L1 interference (ibid. 170), such as the omission of *it* (e.g. **is not*

impossible for language learners to achieve this level) or wrong word order (e.g. **it was developed a pilot assessment tool*); however, their percentage of use seems to decrease from the first to the third year of university, which denotes an improvement over the years.

In the fourth study the focus is on the use of connectors in argumentative essays and takes as a starting point the notion of previous studies by Kaplan that "not only language but also logic and rhetoric are culture specific" (Prat Zagrebelsky 2004: 139) and that in particular Romance languages essays tend to include digressions and additional information that in the English language would be considered superfluous. The analysis examines essays written by Italian students both in English and Italian on the same topic and takes into account the same division of connectors of Quick et al., namely listing (e.g. *first*, *second*...), summative (e.g. i*n conclusion*), appositive (e.g. *for example*), resultive (e.g. *consequently*), contrastive (e.g. *on the contrary*), inferential (e.g. *in other words*) and transitional (e.g. *by the way*). The results show a general

higher frequency of connectors used by learners as compared to natives, and specifically the conjuncts overused by learners are the listing and summative ones, while contrastive and inferential are underused; the category of corroborating ones added by the author (e.g. *as a matter of fact*) also results overused, in particular *in fact*, along with its Italian equivalent *infatti*. Another finding is that while in native writings these connectors tend to occupy a mid-position, in the ones produced by Italians they appear more often at the beginning of sentences. In conclusion, this investigation seems to show the preference of Italian students of a rhetorical model in writing, which means that they value their arguments by providing additional information and examples, while British and Americans rather articulate their arguments through premises, counterarguments and conclusions.

The last study regards argumentative essays too, but takes into consideration the use of direct questions: these are classified as rhetorical, when introduce or change topics (e.g. *what then is ethics?*), and topical, when are used to persuade the reader (e.g. *all acts are equal in the long run, aren't they?*). What could be observed is that while rhetorical questions are not frequent in both corpora, topical ones are more used, especially by learners. Furthermore, while in British texts the interrogative clauses are mostly followed by declarative sentences, in the Italian ones there is a prominence of the personal pronoun *I*, as in expressions like *I think*, *I believe* etc. This shows not only that Italian learners tend

to transfer the same mechanisms of oral exposition also in their writing (a fact that is also proven by the frequent use of *but* as a discourse marker), but also that in order to make their argumentation more effective and persuasive, they tend to rely on personal involvement, which is more a cultural aspect than something that defines their non-nativeness.

Prat Zagrebelsky (2004) also mentions some other statistics which could be collected from the comparison of the ICLE-IT and the LOCNESS corpora: from an analysis of the texts, it can be argued that Italian learners tend to write longer sentences than native speakers, and that their language is less rich and more repetitive in lexical terms. Among the most overused words there are the pronoun *I*, the verb *think* and the adverbs *but* and *moreover*, accordingly to the studies mentioned previously.

As for errors, they are dominant especially in the grammar area and often highlight the influence of the Italian usage of words (e.g. the wrong form *informations* instead of *information*, as it is countable in Italian, or the use of the concessive *even though* as a synonym for the hypothetical *even if*, as in Italian they are rendered by the same conjunction *anche se*). From further research made by students in their dissertations, it could also be shown how learners use a high number of relative constructions and tend to intensify adjectives, but through a limited set of terms (usually through *very*); on the other hand, a lower frequency is found regarding phrasal verbs, and punctuation too is a feature that is often missing.

2.3.2 German learners of English research

A number of studies which analyse native speakers' traits compared with German learners' ones is also accessible: one example is the work of Lorenz (1998), who investigated adjective intensification starting from texts from German learners (both teenagers and university students) and English ones from the LOCNESS corpus.

The first striking result found is that German learners use far more intensifiers than natives, in particular the younger students; then, after observing the actual most intensified adjectives by both groups of learners, the author noticed how seven out of ten were the same, thus drawing the conclusion that intensification has the same function for both, namely to highlight and enhance the meaning of those defined as 'core adjectives', like for example *good*, *difficult*, *important*, etc. In addition, the various categories of

intensifiers proposed by Quirk et al were analysed, which are maximizers (e.g. *absolutely*), boosters (e.g. *very*), approximators (e.g. *nearly*), compromisers (e.g. *rather*), diminishers (e.g. *slightly*) and minimizers (e.g. *scarcely*): the results showed how German learners overuse each single category, meaning that this represents a mechanism which they completely master. Furthermore, the fact that German learners use also many synonyms of those most frequent adjectives (for example, *enormous* and *giant* as for *big*), as well as the device of affixation (e.g. *ultra-big*, *super-important*) indicates that the overuse of intensification is not a strategy to hide any lexical insufficiencies; however, as some times the collocation of particular

adjectives with intensifiers results in expressions which would sound unusual in native writing, learners could be taught to use intensification only when necessary, as "not every noun needs to be premodified by an adjective, and not every adjective in turn needs to be intensified" (Lorenz 1998: 64).

Wh-clefts have been investigated also in German learners in Callies' study (2009), who considers also the Polish ones and uses the two components of the ICLE as corpus reference together with the LOCNESS. The results show how, as well as Italian learners, also these two categories of learners overuse the structure if compared to natives: in specific, learners tend to use them more for emotive evaluations (e.g. *what I enjoyed the most from the beginning to the end was that everybody could wear whatever he wanted*) rather than for highlighting or summarising an action, which is more found in the native corpus (e.g. *what the teams do is the managers get together and decide if they want you to play with them the next year*) (Callies 2009: 289). This indicates a more subjective style of writing of learners, which can be seen also in their significant use of personal pronouns, comparative forms and intensifying adverbs (e.g. *what is even worse is that I'm the only daughter of my father*) (ibid. 290).

Another study by Rogatcheva (2012) focuses on the tense-aspect system and the differences between the English and German one, considering the Bulgarian one too. The author explains how verb forms in the English language are not only numerous, but particularly difficult to learn as well, as they express temporality in different ways and "carry a wide range of semantic meanings such as tense, aspect, mood, modality and concord, and are crucial to the acquisition of the English verb phrase" (Rogatcheva 2012: 259). The major tense constructions identified are present and past, while the major aspect

ones are the progressive and the perfect: the results show how the area less prone to errors is the present, although it is often overused, while the past tense and the progressive and perfect aspects are often misused in both corpora. In particular, the perfect seems the most difficult for German learners, since students do not use it in required contexts and often interchange it with the simple past, not differentiating the two different functions of indefinite (perfect) and definite (simple) past. The comparison with the Bulgarian subcorpus shows how the two categories of learners present different difficulties, as the latter produce more errors using the progressive aspect. This could lead to think that the reason may be different native language backgrounds, however, as the German language presents the simple past and the perfect as well, it is probably caused by not enough focus in the school environment: further and differentiated exercises should therefore be provided to the two categories of students.

Aijmer (2002) investigates modal words (meaning not only verbs, but other devices such as adverbials too), by comparing the German sub-corpus of the ICLE with the Swedish and the French ones. As for modal verbs, they result to be overused by all three categories of learners, especially *might*, while *can* and *could* are more used in particular by the German ones: this fact can be explained by a more informal register used by learners, even in academic writing, and by the large amount of attention dedicated to modal verbs in many textbooks, which means that learners are not aware of the alternatives to express the same concepts. In fact, through a comparison with the native corpus, it is possible to see how many more different patterns and combinations are present in the native language (e.g. *might perhaps*, *may conceivably*, etc), while in the non-native corpora any collocations of this type are more typical of a spoken register (e.g. *probably should*, *obviously can*, etc). Moreover, as one of the main functions of modal elements is to express personal opinions by providing arguments, their overuse in non-native essays results in a more direct and emphatic style.

The German sub-corpus of the ICLE (GeCLE) is used also in the work of Nesselhauf (2005) about the use of collocations: she highlights how they represent a major aspect to learn in a language, since through previous corpus studies has been possible to determinate that among recurrent patterns, over a third of these combinations are actually collocations, in both spoken and written productions. Her analysis takes into account written argumentative essays and is restricted to the verb-noun collocation type (e.g. *do*

a favour), as they tend to be the more numerous and difficult to learn for students. In particular, these are classified into free combinations, where all the elements are used in a literal sense (e.g. *read the paper*), restricted collocations, where at least one element has a non-literal meaning and is restricted to some other certain ones (e.g. *make a decision*), and idioms, which have a figurative meaning (e.g. *kick the bucket*); moreover, not only the lexical items that form the collocation are counted, but also the other elements closely associated with them, like prepositions or pronouns (e.g. *put pressure on somebody*). The results show that many collocations are overused by the learners, meaning that they occur more often than in the native corpus considered (the British National Corpus), such as *solve a problem*, *take care of*, *have a look* etc, probably because they are considered familiar and safe to be used, the so-called "lexical teddy bears"; on the other hand, one third of all the collocations produced unacceptable or questionable results, as another one would have been more appropriate or as they are a word-for-word translation of a collocation existing in the learners' language.

Specifically, learners tend to have more difficulties with the production of verbs, as in most cases their usage is arbitrarily restricted: they can in fact be inappropriate, especially in the case of high-frequency verbs (e.g. *make an experience* instead of *have an experience*), they can be inappropriate only in a certain context but acceptable in others, or be used with the wrong preposition (e.g. *have a look on* instead of *have a look at*), and sometimes even both, as it was found that among the 748 deviant collocations produced, 86 have more than one wrong element. In some cases, learners also used phrasal verbs where a simple verb was more suitable (e.g. *go back to crisis* instead of *have* (*another*) *crisis*), or the verb was simply redundant (e.g. *for that reason, you would not think twice about a separation because it is easy* (*to do*)).

After the verb, the noun is the other element more often misused: the inappropriateness can regard the number, as uncountable items can be produced as countable (e.g. *troubles* for the uncountable *trouble*), the non-existence (e.g. *divorcion* instead of *divorce*) or the choice of the simple noun (e.g. *relation* instead of *relationship*); compounds can also be erroneous or used instead of simple nouns and vice versa (e.g. *telephone numbers* instead of *hotline*, *commercial blocks* instead of *commercials*, *adventures* instead of *adventure stories*). As well as for verbs, the author remarks how learners have most problems with common rather than rare nouns.

Determiners are another complicated category to use for learners: in fact they often produce superfluous articles (e.g. *get (the) permission to* + infinite), on the other hand they do not produce them at all (e.g. *run risk of* instead of *run the risk of*), or articles can be misused and confused with a possessive pronoun (e.g. *pollute our environment* instead of *pollute the environment*).

Noun complementation generates some but less errors, above all through the usage of prepositional phrases instead of clauses (e.g. *I got the impression of...* instead of *I got the impression that...*) or confusion of *to*+infinite clauses and *-ing* ones (e.g. *have a right of -ing* instead of *have a right to*+infinite).

In other cases, it is the whole collocation that needs to be replaced, and not just some elements of it: some examples show the tendency of learners to prefer a stretched verb construction instead of an 'unstretched' one, meaning a derivationally related structure, that would sound more suitable in a particular context (e.g. *give the advice* instead of *advise*); some others show how sometimes the learners use a verb-noun combination to convey a concept which would have been better rendered by a different expression (e.g. *train my muscles* instead of *exercise*).

Regarding the amount of errors in the study, the author points out a difference between deviations in collocations and collocational deviations, where the latter consists in an inappropriate choice of a noun if the verb or the noun could be changed to make the combination acceptable (Nesselhauf, 2005: 126), while the former may regard for example the use of a free combination instead of a more appropriate collocation.

In order to explain these difficulties which even advanced learners go through, she argues that collocations are probably not rooted enough in the learners' competence to be actually used correctly: a phenomenon frequently encountered is in fact the use of blends, which are combinations that present parts of similar expressions, produced in the assumption that they are equally acceptable (e.g. *take care for* as a blend from *take care of* and *care for*). Furthermore, some errors are considered to be L1-induced (e.g. *to win somebody's interest* instead of *arouse somebody's interest* from the German equivalent *jemandes Interesse gewinnen*), which can also be seen in the inappropriate translation out of several possible ones for an item (for example, the German word *Möglichkeit* can be rendered as *possibility*, *chance*, and *opportunity*). More focus should therefore be posed to this area starting from what students confuse more frequently, for example through exercises that

contrast similar collocations in typical contexts simultaneously, in order to enable them to distinguish the many different meanings and usages.

Wiemeyer (2017) investigates the strategies and difficulties in the use of paraphrases by German learners by taking into consideration the Corpus of Academic Learner English (CALE). The texts used were reading reports of over 6000 words, namely summaries of academic research articles, and the students, in their second year of University, were asked in particular to report the authors' questions and hypotheses, the main points and the general argumentation using a maximum of 600 words. In the investigation, the productions of learners were classified in the categories of exactly copied (no words were modified), slight modification (some words and phrases were replaced by synonyms), close paraphrase (the syntax was altered) and total paraphrase (sentences were reformulated and restructured with some words taken by the source text), a taxonomy used in previous studies too. The results showed how direct quotes had not been used frequently, and exact copies were not used at all, while a much higher use of paraphrases was found, in particular of the close type (more than 60%) followed by the total one (around 30%). It could therefore be observed how German learners were able to paraphrase texts by changing both grammatical and lexical items. However, a more detailed look at the data revealed some other aspects still to be mastered: the majority of students in fact did not provide a reference while paraphrasing, but only with direct quotes, and when reporting these, they often used the format "author's name + reporting verb + *that*", with no stylistic variation, probably due to a lack of instruction of some other alternatives.

2.3.3 Spoken learner corpora

All the studies mentioned above regard written learner corpora, but research has been carried out on spoken ones too: the spoken language can be considered in fact "a language variety in its own right" (Mukherjee 2009: 204), as it presents many differences compared to the written one regarding both lexis and grammar, and for this reason the author argues that forms and structures of spoken English should actually be taught as much as the others of written English.

As Prat Zagrebelsky (2004) explains, the project of collecting spoken data is called LINDSEI (Louvain International Database of Spoken English Interlanguage) and, as well

as the ICLE, it collects learner interlanguage produced by various students who learn English as a foreign language, which can be compared to the spoken English produced by natives. A main difference that regards also written and spoken corpora in general, is that the components of the LINDSEI are smaller than the ones of the ICLE: it is in fact much more difficult and demanding to record and transcribe spoken data as compared to the transfer of written ones.

The spoken data in question are above all informal interviews between teachers and intermediate EFL students, and regard in particular conversations of about fifteen minutes about a topic they can choose, followed by some personal questions from the interviewers; for this reason, the studies involve linguistic features typical of this kind of spoken discourse. An analysis of these interviews showed how the amount of utterances of teachers is much smaller than the students' ones and is formed basically of words that the students may be short of in some specific moments and other expressions to prompt answers or comments or to make the discourse more fluent. As for students, it is important to gather more data, as they are the focus of the research: their productions have been found to present many repetitions and overlaps, as they have to improvise what to say and do not have the time to rearrange their sentences the way they can do on written essays.

The author mentions two studies which took into account specifically the Italian sub-corpus of the LINDSEI (LINDSEI-IT). The first one aims to analyse the differences between the spoken and the written communication of learners in terms of vocabulary, thus comparing it with the ICLE-IT; the results show how spoken learner English is characterised by shorter words, generally associated with an informal register. Furthermore, referring to the type/token ratio, namely the number of lexical items in a text, the percentage results higher in the written corpus, which thus presents a richer vocabulary; written texts present also a higher lexical density, which is the proportion of lexical items compared to the number of grammatical items.

The second study takes into consideration only the LINDSEI-IT and analyses discourse markers, as features used to structure conversation in terms of cohesion and coherence. These can be of many types: non-verbal signals like gestures, sounds (e.g. *mmm*), interjections (e.g. *oh*), conjunctions (e.g. *and*), adverbs (e.g. *actually*) or clauses (e.g. *I mean*); as for their functions, the two types identified are interactional, which means that they are used to control cooperation between speakers, and textual, which aims to

structure the text to get more cohesion as possible, as often it results in a difficult procedure having no time to plan it.

Discourse markers result to be more used by the teachers who took the role of interviewers, while a low frequency is found in students' productions. However, an interesting fact which has been found is the frequency of *I mean* and *you know*, which shows how students are often insecure about their point of view and thus need to rephrase their discourse.

Castello (2013) also investigates the use of discourse markers by Italian native speakers, and analyses in particular the University context: the small corpus in question is made up of Skype conversations recorded with the software *Pamela*, and the participants are Italian students of 'Linguistic and Cultural Mediation' at the University of Padua, with a level of proficiency between B1 to B2. Through a comparison with the native sub-corpus from the International Corpus of English – Great Britain (ICE-GB), it was possible to confirm the data from the study previously mentioned, namely that Italian students underuse discourse markers, in particular smallwords like *right* or *well*, while they rely more on pauses and fillers like *ok* or *yes*. Moreover, a L1-induced error that learners often make regarding *yes* is its use at the beginning of a new turn to take the floor, as the equivalent *sì* is used in the Italian language with the same function.

The data show therefore how their fluency needs to be developed through more work, especially through data-driven materials that could be useful for showing the smallwords more typical of the English language.

The German sub-corpus of the LINDSEI (LINDSEI-GE) is investigated instead in the study of Götz (2012) about the concept of fluency, as compared to the Louvain Corpus of Native English Conversation (LOCNEC). As explained by the author, fluency is an essential aspect of oral production, especially its temporal variables like speech rate, length of runs, and number and position of filled and unfilled pauses.

Speech rate does not refer to the literal speed of speaking, as the author explains, but includes also articulation rate and pause time and it is measured by calculating the raw number of words per minute, including self-corrections, repetitions and other hesitation features; the results show how learners produce fewer words than natives, which can thus explain not only their tendency to speak slower, but also a wider use of pauses.

The length of runs consists on the other hand in the amount of speech between hesitations or pauses, and is calculated considering the words uttered until the end of a turn or until

a pause, including repetitions and any interrupted ones; the results show shorter runs produced by learners as compared to natives, to the point that "even the longest MLR (mean length of runs) of 8.91 in the learner corpus is still significantly shorter than the LOCNEC mean" (Götz 2012: 235).

As for pauses, the unfilled ones (UPs) can be divided based on their length, but as the author suggests, since they do not have a significant impact on learners' fluency, no distinctions are taken into account in this study. Overall, UPs are overused by learners, and moreover, through a further classification based on their position, unfilled pauses within clauses, counted as hesitation pauses, result overused not only by learners as a whole, but also for each individual speaker: to explain this in Mukherjee's words (2009: 222), "advanced learners, unlike native speakers, have to consider much more thoroughly not only what they want to say but also how they should say it even within syntactic constituents". An interesting fact regards filled pauses (FPs) instead, as although the majority of learners overused also these ones (80%), a percentage actually used them less than native speakers (8%), meaning that this does not represent a difficult area for all German learners. In order to bridge the gap between learners and native speakers, the author mentions how previous research documented an increase in learners' fluency after a period abroad in an English-speaking country; however, this should be complemented already in the classroom through the use of concordance lines that show the variety of other methods and native-like strategies to improve performances, as for example the use of discourse markers. Recently, many new corpora have been compiled from which concordances of native contexts can be extracted: in particular, the author cites the Corpus of American Television Series (CATS), which can also be found more motivating by students.

Mukherjee (2009) investigates verb-noun collocations produced by German learners comparing the LINDSEI-GE with the GeCLE sub-corpus, thus complementing the study of Nesselhauf (2005): he observes how the most frequent collocations in GeCLE do not appear at all in LINDSEI-GE, which demonstrates that learners distinguish features of spontaneous speech and of essay writing. An additional result is that the whole range of collocations is lower in the spoken language, but it is used much more frequently if compared to the one in written language, as the former tends to be less varied and not proof-read as the latter.

Another study by De Cock, Granger, Leech and McEnery (1998) focuses on the phrasicon, which can be defined as the study of the many ready-made expressions used in the spoken discourse, which enable learners to speak without too many pauses or hesitation. The authors consider in particular learners of French mother tongue, and compare the data with L1 speech; as for the formulae examined, they consist specifically of two to five words that occur with a frequency greater than 9, 4, 3 and 2 respectively. What could be observed is that longer word combinations are significantly more frequent in the non-native corpus, while for shorter ones the data tend not to vary consistently: this shows how learners do use prefabs, and since they also tend to repeat some more often than natives, this fact can confirm how they appear to use them habitually. Furthermore, a deeper analysis of all formulae shows how they often express vagueness (e.g. *and things like that*, *or whatever*, *sort of*) in both corpora, however with a higher frequency for natives: learners tend to use them especially in combination with noun phrases (often to hide gaps in vocabulary, as in *a sort of vapeur*), which is more characteristic of the written register. The authors explain these data by giving a range of possible causes, among which the differences between the mother tongues and the lack of contact with native speakers.
In conclusion, Jones, Byrne and Halenko (2018) analyse in their work strategic and discourse competence of students based on studies from various spoken learner corpora: strategic competence refers to the techniques through which speakers face an obstacle in conveying their meaning and the devices through which they compensate for deficiencies in the target language system. The most used production strategies result correction and utterance reformulation, as well as interaction with the other speakers, for example through follow-up questions in order to seek clarification (e.g. *could you repeat that?*).
Discourse competence on the other hand "concerns the selection, sequencing, and arrangement of words, structures, sentences and utterances to achieve a unified spoken or written text" (Celce-Murcia, Dornyei and Thurell 1995 in Jones, Byrne and Halenko 2018: 110), and thus in speech production refers also to the ability to continue a conversation and to respect turns. The authors show how learners use massively cohesive devices like *this* and *that* but also like *what do you think?*, in particular the most advanced ones, and how they are therefore able to construct turns and link their thoughts to the other speakers' ones developing a coherent and cohesive discourse.

Learner corpora can thus allow researchers, as well as teachers and students themselves, to investigate a wide number of aspects in the production of learners of a language, to compare them with other L1 students and natives and to carry out studies over the years in order to denote any change or improvement. The studies examined in this chapter could show some features of Italian and German learners and some similarities between the two categories, namely the overuse of cleft sentences and of personal pronouns.

Before my own research about how they use *get*, the uses of the verb and previous research about it will be analysed in the following chapter.

CHAPTER 3

3. THE VERB *GET* AND ITS RELEVANCE IN THE ENGLISH LANGUAGE

Before my own research on the verb *get* in the final chapter, in the current chapter I provide a description of its most peculiar characteristics: in particular, as it is one of the most important high-frequency verbs and its correlations with other items form many phrasal verbs and idioms, a section (3.2) is dedicated to their definitions, a list of the most relevant ones and previous research on them.

3.1 The verb *get*: the main uses and meanings

Among the many verbs in the English language that can be regarded as high-frequency verbs, in the sense that they occur significantly more than others, *get* is one of the main ones, as it is in fact ranked 8th in the list of the most frequent verbs in the BNC (Leech et al. 2001); this is due to the fact that it is particularly versatile, as it is used in a wide range of grammatical patterns and as it can have many lexical meanings (Ringbom 1998: 44).
From a grammatical perspective, *get* in its simple past form *got* forms the semi-modal verb *have got to*, that is commonly contracted to *gotta* in an informal register (e.g. *I **have got to (gotta)** go down there and work on school work all day*); in addition, it can also be used as an auxiliary verb in the formation of the passive voice, in combination with a past participle (e.g. *did it really **get** [blown off]?*).
As illustrated in Biber et al. (1999), the major senses in which *get* can be found are those of obtaining something (e.g. *how much are you **getting** a pay raise for?*) and of moving to or away from something (e.g. ***get** in the car*), which denote activities; in addition, it can refer to a mental state, in the sense of understanding something (e.g. *do you **get** it?*) or of changing from one state to another (e.g. *she's **getting** ever so grubby looking now*). Furthermore, it can also be a causative verb, as it can mean to cause something to move (e.g. *Jessie **get** your big bum here*) or to cause something to happen (e.g. *it **gets** people talking again, right*). The fact that all the examples mentioned are taken from

conversations is not accidental: *get* is in fact in conversation "the single most common lexical verb in any one register" (Biber et al. 1999: 376).

Similarly, Gronemeyer (1999) identifies more in detail the main meanings of *get* as in possession, movement, causation and obligation, while adding the one of permission; as far as grammar is concerned, the inchoative and progressive aspects are also listed, along with the passive one.

In describing the sense of possession, the author remarks how *get* is followed by a noun phrase (NP) and this can be interpreted actively as a synonym of 'obtain', but also passively as a synonym of 'receive' (e.g. *but ask coach Darrell Royal what position he plays and you'll* ***get*** *the quick response, `place-kicker'*), with the subject being a recipient and not an agent; on the other hand, when *get* is used metaphorically in its cognitive sense, the subject becomes an experiencer, and in its causative uses it has the function of causer. Furthermore, possession can also be stative and be expressed by *have got + NP*, with the frequent omission of *have* (e.g. *I'****ve got*** *a new book*; *you see... we all* ***got*** *schedules, like any business!*).

When used in the sense of movement, the verb forms the locative construction *get + NP + PP* (particle or adverb): in particular, it can be constructed intransitively or transitively and thus the movement can refer respectively to the subject (e.g. *and Paul Lipson, as Morris, the faithful one who never* ***gets*** *home to his Shirley's dinner, was fine, too*) or to the object (e.g. *the board would cooperate so far as possible to* ***get*** *the children to where the parents wanted them to go*). In addition, in colloquial speech movement can also be expressed by *get* alone without any complement (e.g. *God dammit,* ***get****. And don't come back*).

Causation is rendered by *get* followed by an object and another argument, which can be an *-ing* participle (e.g. *John* ***got*** *his students* ***working*** *on another topic*), a *to*-infinitive (e.g. *John* ***got*** *his students* ***to work*** *on another topic*), a past participle (e.g. *John* ***got*** *his girlfriend* ***invited*** *to all the meetings*), an adjective (e.g. *John* ***got*** *his feet* ***wet***), a prepositional phrase (e.g. *John* ***got*** *his girlfriend* ***in trouble***), as well as a noun phrase (e.g. *John* ***got*** *her* ***dozens of roses***). This form is referred to by the author as ditransitive, as it presents a double object, and the subject can be either an agent, if it is volitional, or a causer if not.

The form *have got + to-infinitive*, which renders *get* a semi-modal verb as mentioned

before, can have both a deontic function and express obligation (e.g. *we all **got to** put up with inconvenience sometimes*), but also probability with an epistemic function (e.g. *this **has got to** be some kind of local phenomenon*), even if significantly less frequently; when *get* only is followed by *to-infinitive*, it denotes permission instead, as a synonym of 'manage to' (e.g. *you did what you were told and if you did it for long enough you'd **get** to turn around someday and tell someone else what to do*).

When its function is of a change-of-state marker, *get* is followed by an adjective or an adjectival past participle (e.g. *we simply can't afford to **get** Ken mad at us*) and this use is called inchoative. When the past participle is not adjectival but verbal, *get* forms a passive construction (e.g. *don't **get** yourself killed for something that doesn't concern you*): Gronemeyer (1999: 6) explains that it is often difficult to distinguish between the two, and suggests the following criterion: "a thematic relation exists between the matrix subject and the internal argument position of the passive participle (*the conference got started*), whereas in the inchoative the matrix subject controls the external argument of the adjectival participle (*we got started*)".

A final meaning considered by the author is the ingressive one, which is represented by the form with dynamic verbs *get + NP + to-V-ing* (e.g. *that night the older men **got to talking** about going possum-hunting on a moonlight night*) or the one with stative verbs *get + NP + to-infinitive* (e.g. *John **got** the students **to understand** the problem*) and whose function is to denote the beginning of a state; as the two examples show, the construction can be both intransitive and transitive.

Therefore it can be stated that *get* possesses a very flexible configuration which explains its polisemy, which "is in general used to refer to the case where the 'same' word (lemma) is used with multiple meanings that are somehow related" (Viberg, 2002: 120): according to the argument structure linked to it, the meaning of the verb in fact changes.

This peculiarity of the verb is highlighted also in Quirk et al. (1985: 720): the authors in fact use the example of *get* to explain how "a given verb can belong, in its various senses, to a number of different classes, and hence enter into a number of different clause types". Gronemeyer (1999), by taking into consideration the Brown Corpus, provides data about the frequency of all the various meanings of the lexeme too: she finds out that the most frequent use that of movement (37%); causation is also rather frequent (15%), while the inchoative and possession uses are the less present (10% and 7% respectively).

In relation to the passive function of *get*, O' Keeffe, McCarthy and Carter (2007) analyse in more detail its use as an alternative to the standard stative be-passive. From the evidence of the corpus of spoken British English CANCODE, they point out that *get*-passives are more frequently used in spoken than in written English and, as regards their semantic prosody, *get* seems to be used more often for unfortunate, or at least problematic, rather than fortunate consequences for the subject, and for this reason it can be referred to as an attitudinal marker. This is confirmed by previous studies on the subject and by corpus data, as some of the most frequently found phrases are in fact *get arrested*, *get killed*, *get beaten*, *get criticised* etc, which characterize adversative contexts.

As O' Keeffe et al. remark (2007: 107), the use of the standard passive is more frequent as many sentences would sound unusual with a *get*-passive: in particular, those presenting factual information statements (e.g. *the steam engine was invented in the nineteenth century* instead of *the steam engine got invented in the nineteenth century*) and truly stative passives (e.g. *the house is surrounded by fields* instead of *the house gets surrounded by fields*). The cases in which a *get*-passive is preferred is in sentences that put emphasis on a change of state: taking into consideration the phrase *the tape seems to* ***have got*** *stuck*, compared to the alternative *the tape seems to* ***be*** *stuck*, it is possible to note how the first one better reproduces the change that occurred. This is also claimed by Gronemeyer (1999: 12), when analysing the differences between *get* and inchoative verbs: "*get* does not denote the actual process of transition from one state to another as a true inchoative does (e.g. *become*), but rather the initial stages of a state, i.e. the emergence of a new state. This of course implies that a change-of-state has occurred".

On a more syntactical note, a comparison between the two passive forms showed how adverbials can be used in both structures with an intensifying or focusing role, however they never occur between *get* and the past participle (e.g. *I* ***nearly got picked*** *on, but I didn't say yes or no*), while it can happen frequently with *be*-passives (e.g. *it* ***was actually destroyed***). Furthermore, another characteristic that distinguishes the *get*-passive from the *be* one is the focus on the subject: in fact, in a sentence like *he* ***got*** *killed* the main focus is on '*he*' (the patient) and on the action, but not on the agent, while a construction like *he* ***was*** *killed* is more generally followed by the preposition *by* and the agent, who is as important as the event and the patient. (O'Keeffe et al. 2007: 107).

Quirk et al. (1985: 161) also remark how the *get*-passive is more common in informal English and how it puts emphasis on the subject and on what happens to it, adding that often not only is the agent not highlighted, but it is not needed at all (e.g. *I have to* ***get dressed*** *before eight o'clock*; *I don't want to* ***get mixed up*** *with the police again)*: the authors call these 'pseudo-passives', as they do not show the possibility to be expanded by an agent.

3.2 The verb *get*: its many patterns in phrasal verbs and idioms

As well as the various constructions that it can form according to the types of phrases which follow it, *get* also generates multiple phrasal verbs if linked to different prepositions, thus expanding even more the range of meanings that it can have.

3.2.1 A definition of phrasal verbs

Phrasal verbs are structures formed by verbs plus prepositions or adverbs (which are thus denominated particles), that function as a single unit, whose meaning often differs from the one of the main verb: for this reason, they represent a rather difficult area of the English language for learners. As Sinclair (1991: 68) puts it,

> the co-occurrence of two quite common little words can unexpectedly create a fairly subtle new meaning that does not seem to be systematically related to either or both of the original words. The disposition of the words involved, and their syntax, is governed by complex and unpredictable rules.

In the English language it is possible to find verbs that collocate with prepositions but which are not phrasal verbs, the so-called prepositional verbs: the differences between the two are both syntactic and phonological, as stated by Quirk et al. (1985: 1167): first of all, the preposition in phrasal verbs can occur both before and after the noun (e.g. *they called up the dean = they called the dean up*), while in prepositional verbs it must always precede the noun phrase (e.g. *they called on the dean ≠ *they called the dean on*). In addition, only with prepositional verbs can adverbs be inserted between verbs and particles (e.g. *they called angrily on the dean*; **they called angrily up the dean*), as can relative pronouns (e.g. *the man on whom they called*; **the man up whom they called*) and

interrogative words (e.g. *on which man did they call?; *up which man did they call?*); furthermore, in prepositional verbs the stress is usually on the verb, while in phrasal verbs it is on the particle.

As said before, the particles in phrasal verbs can be prepositions (e.g. *look **at** these pictures*) or adverbs, mostly of the spatial type (e.g. *one of my papers has gone **astray***), however many particles can be either prepositions or spatial adverbs, such as *about*, *across*, *by*, *down*, *in*, *on*, *off*, *out*, *up*, etc. (Quirk et al. 1985: 1151).

The other most important feature of phrasal verbs is that the verbs and the particle behave as a single unit and have a meaning of their own (e.g. *give in* which means *surrender*), and they are thus different from other free combinations of verbs and particles, where the meaning is still correlated with the main lexical verb and can be thus predicted from the single meanings of its components (e.g. *he walked past*); for this reason, in free combinations the verbs and the particles can be replaced by synonyms, while this procedure is not possible for phrasal verbs, where the elements have a more fixed structure. Moreover, phrasal verbs can be both transitive (e.g. *someone **turned on** the light*) and intransitive (e.g. *the plane has now **taken off***), and many combinations can be found in both structures (e.g. *the tank blew up*; *they may have blown up the bridge*).

Bolinger (1971) observes how phrasal verbs present differences of register too: although being typical of informal English, some examples are more formal, such as the phrases *he set alight the stars* or *hold high the banners*; another characteristic that these verbs have is that they denote actions and results and have thus an aspectual meaning, and for this reason phrasal verbs formed by stative verbs like *know* or *hope* do not exist. The author also outlines how many words of the English language derive from phrasal verbs, like *makeup*, *dropout* or *spin-off*: this can be explained by the fact that they are much used by native speakers, as they present familiar and manageable elements: it is therefore important for learners to be more exposed to them in their studies of the language, in order to feel confident in their use in a similar way.

3.2.2 Phrasal verbs with *get*

Get represents one of the most important verbs with reference to phrasal verbs, as it forms combinations with all the main particles: Cullen et al. (2000), in fact, take *get* as an example when explaining the concept of phrasal verb, along with other frequent verbs

such as *make*, *give* and *see*. The authors list phrasal verbs along with their meaning, their register and examples taken from corpus material to show how they are actually used. The following are those listed by Cullen et al. (2003: 26) in relation to *get*:

Phrasal verb	**Meanings**	**Examples**
get about	1) move around or travel to different places, the ability to walk or move 2) used when information is spread among people	"he's lost the use of his legs and doesn't get about much any more" "keep this information to yourself; I wouldn't want it to get about"
get across	1) manage to cross an obstacle 2) succeed in making other people understand an idea or a message	"how are we going to get across the river when there's no bridge and we don't have a boat?" "it's up to us to get across the message that, once again, the government has got it wrong"
get along	1) have a friendly relationship with people	"somehow I never seemed to be able to get along with my parents-in-law"
get around	1) move around or travel to different places, the ability to walk or move (same as 'get about') 2) used when information is spread among people (same as 'get about')	"he's lost the use of his legs and doesn't get around much any more" "keep this information to yourself; I wouldn't want it to get around" "Steven eventually got

	3) do something, especially after a difficulty or delay 4) avoid a problem rather than dealing with it directly 5) persuade people to do something or allow to do something	around to painting the garden gate" "we'll just have to face it; I can't see any way of getting around it" "don't worry about Mum; I can easily get around her"
get at	1) ask what people are suggesting or what they mean 2) criticize or find fault with someone	"do you know what he was getting at when he said there was trouble on the horizon?" "I'm fed up with that teacher, she's always getting at me"
get away	1) escape 2) leave 3) go away from home on a holiday	"the police managed to catch three of the robbers but the leader of the gang got away" "I'll have to get away by 11 o'clock if I'm to catch the last train" "we're hoping to get away for a couple of weeks in the summer"
get away with	1) escape taking something with you	"the robbers had got away with nearly £100,000 worth of diamonds"

	2) manage to avoid being caught or punished for something illegal or dishonest 3) manage to do something that others would normally find unacceptable without being found out	"he's apparently been getting away with tax fraud for years" "I sang the first verse twice, but no one seems to have noticed so I think I got away with it"
get back	1) return home or to the place where everything started out from 2) return someone or something home or to the place that they came from 3) order someone to move to a position further back or to move away 4) get something that was lost and belonged to someone 5) used metaphorically as a synonym of 'recover'	"our train was late and we didn't get back until after midnight" "do you think we should take him back for a second interview?" "get back! I've got a gun and I'll use it if you come any closer!" "I lent him my Bon Jovi tapes months ago and I don't think I'm ever going to get them back" "after looking so pale and ill for so long, she's now got her normal healthy glow back"
get behind	1) do not keep up to date with something, for example work 2) do not pay something at the times it is due, for example rent	"this delay was due to one or other of the works departments getting behind with the manufacture" "because the sums involved were so much larger, getting

	3) give support and encouragement to someone	behind with mortgage payments became a much more serious debt trap than bills at retailers" "the supporters should get behind their team instead of continually criticizing them"
get down	1) move down to a lower position from a high one, or from above the ground as a synonym of 'descend'	"the cat has climbed to the top of the tree and can't get down again"
	2) record something by writing it on paper	"the professor talks so quickly those of us who can't do shorthand find it difficult to get everything down"
	3) make people sad or depressed	"his constant criticisms are beginning to get me down"
get in	1) be elected to power, referred to a politician or a political party	"do you think the Tories will get in again at the next election?"
	2) find the time or opportunity for something	"it was difficult to get a comment in, everyone was talking so furiously"
	3) get someone to come to a place of work to do work of some kind	"-will you be putting the new bathroom in yourselves? -No, we're getting a local firm of plumbers in"
	4) arrive at its destination, referred to trains or buses	"the last train gets in at midnight"

get off	1) move off or away from something	"get off my land immediately or I'll have you prosecuted"
	2) remove something	"I'll never get this stain off the carpet"
	3) get out of a train, a bus or a plane	"have my seat. I'm getting off at the next station"
	4) dismount from a horse or a bike	"he had to get off his bike and push it up the hill"
	5) leave, for example to go on a journey	"we've packed all our camping equipment and are hoping to get off as soon as it's light"
	6) send a letter or a message by post or e-mail	"can you make sure this letter gets off before the last post tonight?"
	7) be given little or no punishment for something wrong, illegal or dishonest done	"the lawyers got him off with a small fine"
get on	1) have a friendly relationship with someone (same as get along)	"I don't think my son and his wife have been getting on very well recently"
	2) make progress in a career	"it seems that, in order to get on, you have to socialize with the bosses"
	3) become old	"I don't think Dad will be able to come on a walking holiday with us; he's getting on a bit, you know"
	4) make progress, referring to what someone is doing generally	"how's David getting on in his new job? He seems to be getting on fine"
	5) board a train or bus at the start of the journey	"hurry up and get on the bus"
	6) put clothes on	

		"she was trying to get her new jeans on
get on with	1) begin to do something or carry on doing something	"I had better get on with this work if I want to finish it today"
get out	1) leave 2) become known publicly, referred to secrets 3) escape or be allowed to leave a place where someone had been kept prisoner or contained 4) spend time outside home	"get out of my sight, you disgusting child!" "if this gets out, we'll be in real trouble" "when he eventually gets out, he'll find it pretty difficult to adjust to life outside prison" "she's a bit frail now and doesn't get out much any more"
get out of	1) manage to avoid doing something	"I don't want to go to their stupid dinner party but how on Earth am I going to get out of it?"
get over	1) manage to move over an obstacle to the other side 2) recover from an illness, a shock or a disappointment 3) make other people understand a message (same as 'get across')	"I don't think that little pony will get over that big jump" "she's never really got over the busy period in the summer"

get through	1) finish or complete a task 2) use all of an amount or supply of something 3) succeed in contacting someone by telephone 4) make someone understand or realize something	"it'll take me hours to get through all these letters" "we seem to be getting through an awful lot of coffee" "I tried ringing him at the cottage but I couldn't get through" "I can't seem to get through to them how important this is"
get to	1) reach or arrive at a place 2) arrive at the time or place someone was expected 3) upset or annoy someone	"what's the best way to get to the British Museum from here?" "he should have been here ages ago. Have you any idea where he might have got to?" "you shouldn't let him get to you. He's like that to everyone"
get together	1) meet, especially informally	"we must get together soon and discuss the project"
get up	1) leave the bed after waking, or stand from a sitting or laying position as a synonym of 'rise' 2) make someone get out of bed, usually earlier than intended	"everyone got up when she came in the room" "she got me up at six o'clock just to watch the sun rise"

get up to	1) do, referred as what someone thinks someone else is doing, especially if it is likely to be something that will be disapproved of	"I hate to think what he and his mates are getting up to in Greece"

Some of these are more representative of an informal register than others, such as *get around*, *get along* and *get at*; moreover, it is notable how phrasal verbs can also be formed by a verb plus more than one particle, as in *get away with*, *get on with*, *get out of* and *get up to.*

McArthur and Atkins (1974), who provide examples of phrasal verbs with a very short definition, include additional ones regarding *get*:

get above (oneself)	1) pretend to a superior social position	"oh, she really is getting above herself!"
get ahead	1) go well in advance 2) progress well	"that ship has got ahead of the others" "he's certainly getting ahead in his profession"
get back at	1) gain revenge on someone	"she got back at him by showing me the letter"
get by	1) pass 2) be tolerable 3) manage	"let me get by, please" "this work will get by, but that's all" "she gets by on a remarkably small income"

get forward	1) move forward 2) convey forward	“they got forward to the front line without any enemy response” “we got it forward on time”
get into	1) be involved in	“the children are always getting into trouble/mischief”
get in with	1) get into the good graces of someone	“he is trying hard to get in with those people, because he thinks they can help him”
get off with	1) start having an affair with someone	“he got off with that blonde”
get on for	1) come close to something	“he must be getting on for forty now”
get on to	1) recognize 2) nag 3) contact someone	“the police will get on to him very soon” “she’s always getting on to me about something” “I’ll get on to them straight away and find out what is happening”
get round to	1) find time to do something	“I always wanted to write, but have never got round to it”
get through with	1) finish completely, with a feeling of relief	“he finally got through with the subject”

3.2.3 Idioms with *get*

In relation to the concept of phrasal verbs, it is important to mention idioms too: they can be assimilated since they are both not compositional, which means they are groups of words whose overall meaning is different from the meanings of the individual words; nonetheless only phrasal verbs are formed by a base verb and at least one particle, while idioms can present a more varied structure. Gairns and Redman (2011a) note how students should learn them together, as they occur frequently in the English language, allowing thus teachers to explain their similarities and differences.

Examples of idioms provided by Gairns and Redman are taken from various ELT dictionaries (*Oxford Advanced Learner's Dictionary*, *Oxford Wordpower Dictionary*, *Oxford Idioms Dictionary for learners of English*, *Longman Dictionary of Contemporary English*, *Macmillan English Dictionary for Advanced Learners*, *Cambridge Advanced Learner's Dictionary* and *Collins COBUILD Advanced Learner's English Dictionary*), since they note that generally dictionaries tend to not agree as to what is considered an idiom, with some of them classifying these structures into collocations. Moreover, their volume provides, as well as explanations of the main idioms for intermediate learners, extracts from texts and many exercises to make learners practice right away. The idioms covered which regard the verb *get* are the following:

Idiom	**Meaning**	**Example**
get a move on	hurry up	"come on, get a move on!"
(not to) get a word in edgeways	(not) be able to say anything because someone else is talking too much	"my colleague talks non-stop! You can't get a word in edgeways"
get away from it all	go on a holiday to a place where you can relax	"walking in the black forest - where to get away from it all"
get bogged down	become so involved with the details of something	"I know it's important not to get too bogged down, so I

	that you can't make any progress	decided not to study everything but just try and work out which questions might come up and concentrate on those"
get butterflies in your stomach	feel very nervous about doing something	"but on the morning of the flight, I was starting to get butterflies in my stomach"
get cold feet	suddenly become nervous about something planned to do	"do you still want to do this parachute jump or are you getting cold feet?"[2]
get hold (of somebody)	find or make contact with someone by e-mail or phone	"well, I've been in touch with most of them, but I can't get hold of Uncle Nick"
get in the way (of something)	prevent or stop something from happening	"something is getting in the way of your happiness"
get in touch (with somebody)	make contact with somebody, especially after a long time	"why not get in touch with a few old friends and go together?"
get into the habit of doing something	develop a particular habit	"I got into the habit of reading the business pages in the newspapers, but I also spent the holidays going through all

[2] As for some idioms there were no examples provided in the volume, I retrieved them from the following website: https://www.thefreedictionary.com/

		my lecture notes”
get into shape	become physically fitter as a result of taking exercise and eating healthy food	“after she had the baby, she started swimming everyday, to get back into shape”[3]
get in(to) trouble	get into a situation in which you may be punished	“we were always getting into trouble at school for smoking, wearing make-up, or just being lazy”
get nowhere	make no progress, have no success	“even Magnus agreed we were getting nowhere and announced that we were lost”
get off on the wrong foot	make a bad start at a relationship	“at first I thought he was a bit out of touch, and he wasn’t very friendly, but maybe we just got off on the wrong foot”
get off to a (good, bad, better) start	start something (well, badly, better)	“despite losing the toss and playing against the wind in the first half, Camborne United got off to a better start”
get on someone’s nerves	annoy or make someone angry	“I love Brad, but some things about him get on my nerves”

[3] As for some idioms there were no examples provided in the volume, I retrieved them from the following website: https://www.thefreedictionary.com/

get on top (of someone)	be too much for someone to manage or deal with	"work is getting on top of me at the moment"
get on top (of something)	manage to control or deal with something	"I've worked hard to try and get on top of this subject"
get on with (your) life	stop worrying about something that has happened and start living a normal life again	"get rid of your bad habits and get on with your life"
get out and about	go to places where you can meet people	"on the other side there's an old lady living all by herself who doesn't get out and about much"
get rid (of something/someone)	take action in order to be free (of it/them)	"open the door and get rid of the smell"
get (something) off the ground	make (something) start happening successfully	"it took ages to take the project off the ground, but the architects finally came up with a design that everyone liked"
get stuck (on something)	not be able to continue with something because it is too hard	"if I get stuck on a question, I miss it out and go on to an easier one"
get the better (of something/someone)	defeat or be stronger than something/someone	"they have 45 minutes trying to get the better of him on the football field"

get the hang (of something)	learn or begin to understand how to do something	"- I've never used this computer before - it's OK once you get the hang of it"
get the most out of something	get the maximum benefit or pleasure from something	"ten ways to get the most out of your day"
get the upper hand	gain an advantage over someone so that you are in control of a situation	"you and your dog – get the upper hand today!"
get the wrong end of the stick	understand something in the wrong way	"I thought he was making a joke, so I laughed, but I'd got the wrong end of the stick"
get there	achieve something after a period of work or effort	"don't give up. Stick at it and you'll get there in the end"
get to grips (with something)	manage to control or deal with something (same as 'get on top of something')	"we still haven't got to grips with traffic congestion and pollution"
get to know someone	meet someone a number of times and become friends	"I've got a cousin who is getting on for forty, but I only got to know him quite recently"
get your money's worth	get the full value of the money you have spent	"how to get your money's worth when hiring a builder:

		read our factsheet and it won't cost you a penny"
get your own back (on someone)	do something unpleasant to someone in return for something unpleasant they did to you	"I got my own back by hiding her favorite doll"
get your own way	get or do what you want, although other people may want something different	"he always got his own way as a child"

In addition, further meanings of some phrasal verbs mentioned before are also taken into consideration, which are:

get back to someone	1) phone or speak to someone, especially to give a reply	"if you leave your number, he'll get back to you later"
get down (to something)	1) begin to do something and give serious attention to it	"anyway, I'd better get down to some work"
get in	1) be accepted to study at a school or university	"I didn't get in at university when I left school, but I went two years later"
get into	1) start a career in a particular profession	"get into advertising"

get off	1) leave the place where you work at the end of the day	"I get off work early on Fridays"
get (something) out of something	1) take something from inside the place where it normally is 2) get pleasure or benefit from something	"the skin opens and you can get the garlic out easily" "I didn't get much out of the course"
get through	1) reach a good enough standard to pass a test	"get through First Certificate: five practice tests"

The fact that two whole pages are dedicated to the verb (Gairns and Redman 2011a: 134, 135) underlines even more how *get* is considered particularly relevant both in phrasal verbs and idioms. Moreover, another meaning of *get* as in *understand* is also mentioned (Gairns and Redman 2011a: 52), which often refers to jokes (e.g. *sorry, I don't get it*, after not having understood a joke), but is also found in the idiom *get the picture*, which means to understand a situation that has been described (Gairns and Redman 2011b: 92).
A version of the volume aimed at advanced learners has also been produced by the same authors (Gairns and Redman, 2011b), with further expressions associated with *get*:

get away with murder	do whatever you want without being stopped or punished	"the last to be born is the baby and tends to get away with murder"
get beyond a joke	become annoying or worrying	"it got beyond a joke"

get carried away	become so excited or involved in something that you lose control of your feelings or actions	"he gets a bit carried away at times"
get caught up in something	become involved in something which may cause problems	"he lost his job and got caught up in a downward spiral that led to him becoming homeless"
get (your) foot in the door	get your first opportunity to work for an organization or business, which could later bring you success	"first, you need to get your foot in the door"
get (your) hands on something	find or get something that you want or really need	"I could get my hands on about £200 to help out, but he'll need more than that to pay off his debts"
get on like a house on fire	get on well	"the two boys get on like a house on fire"
get out of bed on the wrong side	be in a bad mood (used especially in a humorous tone)	"I think Ann got out of bed on the wrong side this morning"
get out of hand	become difficult or impossible to control	"inflation is getting out of hand"
get soaked to the skin	get very wet	"last year, it tipped down all weekend and we got soaked

		to the skin”
get someone’s back up	annoy someone	“don’t get people’s backs up by being rude, aggressive, dismissive, etc.”
get someone into deep (or hot) water	get someone into a serious or difficult situation	“when we realized we got in deep water, it took us ages to cut our losses and sell up”
get the better of someone	behave in a way that you do not want to, when emotions are too strong to control	“his emotions got the better of him and we couldn’t shut him up”
get the push (or boot/elbow)	be dismissed from a job	“I got the push from my first job in a solicitor’s office”
get to the point	stop talking about unimportant details and say what is most important	“stop beating about the bush and get to the point of what you’re trying to say”
get what (you) deserve	earn the bad things that happen to (you)	“in life, we get what we deserve”
get worked up about something	get very excited, angry or upset about something	“don’t get worked up about not sleeping, it only makes things worse”

As for phrasal verbs, the only one included and not mentioned before is the following:

get round someone	persuade someone to do what you want, often by being nice to them	We'll have to find a way to get round the accountant if we want this tax scheme to work.[1]

3.3 Studies on high-frequency and phrasal verbs

3.3.1 High-frequency verbs research

High-frequency verbs can be referred to as basic, simple verbs that are particularly frequent in all languages: for this reason, they are already "encountered in the early stages of learning English" (Lennon 1996), however, as some authors remark (Gouverneur 2008; Gerckens & Gans 2015), they still tend to be a problematic area for advanced learners. Being the most frequent verbs means presenting many meanings and structures, as seen with *get*: their polysemous character seems to be in fact the main reason for the problems encountered by learners, as they tend to understand and master the core meanings, but to have more difficulties in the production of their phraseological patterns.

The authors mentioned above take the verb *make* into consideration: Gerckens & Gans (2015) consider the German sub-corpus of the LONGDALE project for learner data and the LOCNESS as a reference native corpus, and show how the verb is underused by L2 users of English as compared to native speakers; furthermore, the verb is often misused by students, as another choice would be more appropriate (e.g. *make* instead of *have an experience*), which shows how frequently the negative transfer from the native tongue causes the most mistakes. However, the limited text types (in this case, narrative, argumentative and expository) can also be considered another factor which explains the low frequency of the verb, since learners tend to underuse also causative structures of *make (*e.g. *to make people talk)*, present nonetheless in the German language.

On the other hand, Gouverneur (2008) takes into account the other potential factor of the incomplete design of learners' material and analyses how phraseology is treated in

[1]As for some idioms there were no examples provided in the volume, I retrieved them from the following website: https://www.thefreedictionary.com/

textbooks of intermediate and advanced level (in particular *Cutting Edge*, *Inside Out* and *New Headway*). She refers to previous studies (Sinclair & Renouf 1988, O'Dell 1997 in Gouverneur 2008) that already stated how coursebooks tended to focus more on the lexical meanings instead of the delexicalised ones (mainly phraseological patterns), and then focuses on her own study about *make* and in particular its most problematic 'verb + object' constructions: surprisingly, the highest proportion of exercises regards restricted collocations, which seems to show that phraseology is strongly present in textbooks, however a deeper investigation into the single exercises shows how often collocations are only included in the task but are not its main focus (e.g. exercises to find a synonym for an expression containing *make*, where the attention is not drawn to it), especially in the advanced textbooks. The author points out therefore how proficient learners should enhance the knowledge that they acquire in an intermediate level, but they often cannot through their textbooks as these do not provide the adequate focus for these structures, that are mistakenly considered already mastered.

A different result is found in Ringbom (1998), who considers the most frequent words taken from some ICLE sub-corpora for non-native speakers (among which French, Spanish and German) and the LOCNESS for native ones. Taking into consideration only high-frequency verbs (as many high-frequency items are articles like *the* or prepositions like *and* or *in*), some of them are found in fact more in non-native texts as compared to native ones, especially auxiliaries like *be*/*have* or *can* and others like *think*, *want* and *know*: this can be explained by a massive use from learners of phrases containing these verbs, such as *I think* for expressing personal opinions, but also by the fact that they tend to be used mostly with their main meaning and not to form structures like phrasal verbs (as mentioned in 3.2.1). However, also the verb *get* is among the verbs overused by learners, but as the author explains, it illustrates one of their many shortcomings: in fact they often use this verb when an alternative one would be preferable, showing therefore a limited vocabulary (e.g. *get* instead of *form their own opinion*), which is confirmed also by an overuse of general and vague nouns like *people* and *things* (Ringbom 1998: 49).

The difficulty in using high-frequency verbs for learners is studied also by López Pérez and Benali Taouis (2019), who consider in particular the verb *do* and the Spanish corpus ENTECOR, formed by texts produced in online forums from a subject within the framework of the Degree in Early Childhood Education. After having eliminated the

occurrences of *do* as an auxiliary and as a form to emphasise affirmative sentences (the so-called *emphatic do*), it was possible to trace in the corpus the uses of the verb needed for the study: out of 164 collocations, 42 errors were found, meaning that in many instances the verb is misused (25%). The mistakes encountered regard the use of *do* instead of *make*, as used in the sense of 'create' (e.g. *they would *do the activities according to the needs of each school and children*, instead of *make the activities*) and in delexical structures (e.g.*all the cuts that have been *done in education*, instead of *all the cuts that have been made in education*), and its use when another verb was more adequate (e.g. *it was a pleasure to *do this forum with all of you*, instead of *it was a pleasure to share this forum with all of you*). As possible explanations, the authors refer to negative transfer from L1, as in Spanish *do* and *make* are both rendered by the verb *hacer*, a particular difficulty with delexical structures, that seems to confirm the data found by Gouverneur (2008), and a low vocabulary and above all collocation competence.

These features result to be common to most learners and high-frequency verbs, as also another study by Zhou (2016) that takes into consideration Chinese ones and the verb *have* shows. After having removed the auxiliary forms of the verb as in *do* in the previous study, the data taken from the Chinese Learner English Corpus (CLEC) illustrate in fact the same characteristics: Chinese learners as well tend to produce deviant collocations with the verb *have*, in particular forming expressions that have a correspondence in their native language but sound unnatural in English; furthermore, the author also notes how the correct ones tend to be repeated frequently in the writings, with no use of alternatives and above all with a strong reliance to the words already encountered, for example in the titles of the essays.

All these studies depict how learners, not having enough knowledge of these structures, tend to use high-frequency verbs as they are seen as familiar and safe choices, which can explain their overuse as compared to native speakers; more focus needs to be aimed thus to the various patterns of these verbs and to the differences in the way they are rendered in the English language and in the L1 of the learners. Even though *get* is among the most important and frequent of these verbs, corpus-based research about its use by different learners of English is in fact notably scarce, as only one study could be found (see 3.5).

3.3.2 Phrasal verbs research

Many studies about phrasal verbs, which are often formed by high-frequency verbs and other particles, are also present: as some authors remark (Gardner & Davies 2007, Liu 2011), this category of verbs is difficult to learn for students as well and is therefore not frequently used. This is in particular due to their complex syntactic structures, since some are fixed expressions while others allow variations in the position of the items; in addition, many of them present more than one meaning, thus it is often easier for learners to use single verb alternatives. Moreover, as well as for idioms, there is not a total agreement between experts about what is considered a phrasal verb, and since as a result most sources are not homogeneous, it generates confusion among the learners. What is more, as Liu (2011: 662) notes, "the enormous number of PVs in English also contributes to the problem, because it makes learners feel overwhelmed, not knowing which ones to learn". Therefore it is relevant to define the most useful phrasal verbs for students to learn, which are generally the most highly frequent: corpus-based studies have thus been carried out in order to identify them.

Gardner and Davies (2007) took as a reference the British National Corpus (BNC) and through the WordNet program were also able to make distinctions between the various different senses of the same word forms, in order to avoid any kind of oversimplification. The authors classified the frequencies of preposition-adverbial particles, of lexical verbs and of verb-particle combinations and came up with some interesting findings: there are in fact some particles that can occur both highly frequently with some verbs and never once with others (e.g. the particle *on* with respectively the verbs *go* and *point*), while some other particles can be generally infrequent in the use of phrasal verbs, but nonetheless form a few combinations that result among the most frequent in the corpus (e.g. the particle *round* which usually does not form phrasal verbs, in the combination *go round* occurs however 1366 times). This shows another reason as why these structures are difficult for learners: not only do they have to learn any possible combination, but they also need to understand the ones which are less likely and the ones that do not occur at all having no rule to follow, as these verbs present a random nature. As a possible solution, the authors suggest that learners could start from the most frequent particles, from those look for the corresponding verbs that form most combinations, and then follow the same method starting from the most frequent lexical verbs and looking for the

corresponding particles afterwards, in order to recognise the most frequent ones as a starting point. Only subsequently can they analyse these in all their various meanings and practice through exercises in order to be able not only to recognise, but also to use them productively. Liu (2011) takes a step forward and examines the variety of American English too through the Corpus of Contemporary American English (COCA), which contains, as well as academic writing, fiction, magazine and newspapers sub-corpora, a spoken sub-corpus too: the registers of the most frequent phrasal verbs are thus considered as well.

Using some dictionaries as a search list guide for locating the most relevant phrasal verbs, a comparison with the BNC shows that there is not a significant difference between the two, which also demonstrates how their use has remained stable during time, as the COCA covers a period of time consecutive to the BNC (from 1990). However, some differences in the verbs' meanings could be identified: for example, the author observes how *fill in* appears more frequently in the BNC as in British English it

is also used in the sense of 'filling in a form or document', while in American English the same meaning is generally expressed by *fill out*; another example, on the other hand, is *check out*, which results highly more frequent in the COCA as in American English it is used in the senses of 'paying for things' and 'borrowing items from a library', which are not found in British English.

As for the registers, the results show that overall phrasal verbs are more common in spoken English rather than in the other written styles, above all academic writing; nonetheless, some items have a wide distributional range and appear frequently in all the various registers (e.g. *carry on*, *follow up*, *break down*) and therefore, as the author suggests, they should be the starting point and priority of second language learners. Finally, it is important to note how also a single polysemous phrasal verb can be typical of more than one register through its various meanings, and this category is particularly relevant for learners too: the example mentioned in the study is *make up*, which is more found in academic writing in the sense of 'compose, constitute' (e.g. *women make up 22 % of the rural labor force in Nicaragua*), while in conversation is more frequent in the other senses of 'decide' (e.g. *Secretary Powell can make up his own mind*), 'compensate' (e.g. *the kids make up for their lack in experience with enthusiasm*) and 'fabricate' (e.g. *Melanie made up that story*).

Research about the actual use and difficulties of learners in their production of phrasal verbs has been carried out too: for example, Chen (2013) takes into consideration Chinese learners and compares them with native speakers in a longitudinal study: the students' essays have been collected over their three years of University, while for the native speakers' counterpart the *Louvain Corpus of Native English Essays* (LOCNESS-US) and the *General Studies Corpus* (GS-UK) have been used for an additional comparison between the two variants of English. The author states right away how "mastery of a large repertoire of PVs, together with other types of phraseological units, by native English writers is considered an important difference between their writing and learner writing" (Chen 2013: 90). In fact, learners often tend to use alternatives, such as easier and safer to produce single verbs, and this turns out in a significant underuse of phrasal verbs and in a major gap if compared to natives.

The Chinese learners investigated in this study, who are not familiar with phrasal verbs as the Chinese language does not possess this structure, produce in fact half the phrasal verbs in comparison to natives, in particular to American ones. Although the quantity does not differ much in comparison to British English speakers, there is still a difference in their production, as natives show a better awareness of the most frequent phrasal verbs in their corpus, while Chinese learners seem to know only 60% of the most frequent in their learner corpus. Moreover, a further comparison between Chinese students at different levels of proficiency shows how they do not get better with time: on the contrary, their use decreases from the first to the last year, and this seems to prove another similarity between phrasal verbs and high-frequency verbs, as they both still result difficult for advanced learners.

As mentioned above, Chinese learners are not familiar with the structure of phrasal verbs, since they do not exist in their native language; some authors decided thus to carry out some studies in order to investigate whether the L1 background of learners influences their production. Dagut and Laufer (1985) for example take into consideration Hebrew learners, who like Chinese speakers do not have this structure in their mother tongue. Their investigation was formed of three stages: initially the students were given a number of sentences with a blank space to be filled with one of the verbs given at the end of each phrase, among which one being the correct phrasal verb, another one a correct single verb (a synonym of the phrasal verb in question) and two wrong ones (e.g. *we didn't believe*

that John could ever _ his friends – let down, *solve*, *disappoint*, *carry on*). In the subsequent step, learners had to directly translate the Hebrew verb given at the end of the same sentences of the first exercise with their English equivalent (e.g. *we didn't believe that John could ever _ his friends – leachzev*), while the last task was a memorizing test, as students were given the English sentences both with the correct phrasal verb and with the single verb alternative, along with their Hebrew translation, and after memorizing them, they were given only the Hebrew ones to translate. The results showed a general avoidance phenomenon of phrasal verbs: in fact, out of the 900 possible uses of phrasal verbs (15 sentences per 60 students), only 376 were chosen by the participants in the first multiple-choice stage, which corresponds to a percentage of 42% (against the 67% one shown by native speakers who performed the same task); in the second exercise, 32% of learners translated the Hebrew verb with an English phrasal verb equivalent and only 24% of them could memorize the phrasal verb in the last phase, with a relevant percentage of verbs mistranslated or not translated at all (35%). These findings show how this cannot be explained by a general lack of knowledge of the structure, since the phrasal verbs were actually listed in more than one task and a percentage of students, even if small, used them: when expressing themselves in the English language, the majority of learners simply avoids them and prefers a single verb, which results more familiar as similar to any Hebrew equivalent.

The same procedure was followed by Hulstijn and Marchena (1989) in their study with Dutch learners: since phrasal verbs are present in their native language, they aimed to investigate if this was an actual factor that influenced their use, or if they also avoided to choose them as well as Hebrew learners for other reasons; moreover, learners were divided according to their linguistic level in intermediate and advanced, to see if proficiency played a role or not. The tasks were modified a little, in order to get more evidence of an avoidance, and not an ignorance behavior: particularly in the memorization one, no Dutch equivalents were given, since they could possibly interfere with the memory effect through a translation one, and two columns with the headings 'I know the original word' and 'I can't remember the original verb but it could be' were added, since "in cases of reported uncertainty, a low number of phrasal and a high number of non-phrasal responses could be interpreted as evidence of avoidance behavior" (ibid.: 245). The results showed a substantial difference in comparison to the previous study: most

Dutch learners showed in fact a preference for phrasal verbs in the multiple-choice exercise, even close to the percentage of natives, and they could memorize them in the memorization task too. A relevant difference was also discovered between intermediate and advanced learners: the former gave significantly fewer correct and phrasal answers than the latter, showing therefore a different behavior if compared to Chinese learners. The authors explain how this sort of avoidance manifested itself especially with some particular verbs, namely the ones with a different structure from their L1, but also the ones too similar to their L1 for fear of interference (e.g. *give up* because of the Dutch equivalent *opgeven*), and others with too specific semantic features (e.g. *turn up*, while the single verb *appear* is more generic).

Thus it is possible to state that students' native language does play a part in the production of phrasal verbs, and not only differences but also similarities can affect their behavior; in addition, the level of proficiency and the semantic characteristics of some verbs are equally relevant factors that can influence the learners.

Additional explanations for phrasal verbs avoidance are included in a more recent study by Becker (2014) with Chinese learners, namely the type of verb and the learning context of learners. In fact, as in a similar previous study (Liao and Fukuya, 2004), the two groups of ESL and EFL students participated in the study, both formed by intermediate and advanced learners; however Becker found relevant to also add what is actually taught in the two different types of textbooks in reference to phrasal verbs, as to determine if it is proper to define it a case of avoidance. As to the actual verbs, a distinction between literal and figurative ones was considered useful, where the first have meanings that can somehow be derived from their semantic components (e.g. *take away*), while the second have a metaphorical meaning that cannot be ascertained by the single constituents (e.g. *let down*). Moreover, as far as the procedure is concerned, the same tasks of 'multiple-choice' and 'translation' mentioned in the other studies were used, however another one of 'story re-tell' was added: this consisted in giving the students a story to read which included the phrasal verbs contained in the other two exercises (selected among the most used in conversation according to the 'Longman Grammar of Spoken and Written English' corpus), and then to make them re-tell what they could remember in their own words, in order to see if the verbs were actually produced when not supplied with limited choices. The last stage consisted then in completing a survey not only about demographic

information, but also about the phrasal verbs themselves: in fact, a list with the same ones as in the previous exercises was provided to learners, who had to indicate which ones they knew in order to further determine the ones avoided. As expected, from the results it was possible to claim that the 'story re-tell' task was the one where phrasal verbs were mostly avoided, as learners, when having to produce an oral discourse, relied more on single verbs equivalents (e.g. *quit* instead of *give up*): this shows how learners may have a good knowledge of phrasal verbs and of how to use them when given the contexts, however they still present difficulties in generating them autonomously. Another predictable result regards the type of verbs, as the figurative ones, more opaque and difficult to learn for their idiomatic nature, result less used than the literal ones. With reference to the learning context of students, it could be noticed how ESL ones had higher scores in the production of phrasal verbs in the translation task, as they are generally more exposed to the target language and therefore to its vocabulary, while more surprisingly EFL ones performed better in the story re-tell task, probably since they payed more attention and memorized the verbs, above all the figurative ones, with no modification: the type of task considered results thus another important factor that can affect learners' use or avoidance of phrasal verbs.

As well as factors that can affect avoidance of phrasal verbs, authors have also investigated the ones that influence their knowledge: an example is the work of Schmitt and Redwood (2011), who consider both ESL and EFL learners, of both genders, from 14 countries (among which Italy, Spain, Germany and China) and ranging in age from 14 to 55. A list of 60 phrasal verbs taken from the BNC was analysed, including both some of the most and some of the least frequent ones in order to study the relationship between frequency and learner knowledge: the tests provided were both productive and receptive, as the learners were asked first to produce the phrasal verb themselves having at their disposal a single-item synonym, as in the example *the police s_ u_ roadblocks to stop people driving into the city centre* (*build, erect*) and then to choose between alternatives in a multiple-choice task, as in *the police _ _ roadblocks to stop people driving into the city centre* (A: *set in*, B: *set up*, C: *set on*, D: *set at*). In addition, a questionnaire at the end was added, as to see which aspects of language exposure result actually more effective in the verbs' acquisition. The results showed how learners tend to know more the most frequent phrasal verbs, with some variation regarding the rank order of frequency,

with no difference between the written and spoken register and between genders or ages, and with higher scores in the receptive task; moreover, a comparison with the COCA corpus led to similar results. As for the findings from the questionnaire, it could be stated that extensive reading and watching English films and TV shows have a positive effect in the language proficiency, as learners who read and watch those the most (around 2+ hours per week) knew more phrasal verbs from the tests, while other factors like listening to music or spending time on social networks do not affect it as much, probably as they require less concentration on the vocabulary.

3.4 Studies on the verb *get*

After the section about studies on high-frequency and phrasal verbs in general, I wanted to provide some results from other studies about *get* only, before my own research about German and Italian learners of English. However, to my knowledge there is no research based on learner corpora which analyses how learners use this verb, except for the work of Coto-Villalibre (2016), who takes into consideration Hong Kong students and in particular the participial constructions that *get* can form. The author divides these into five categories according to their level of 'passiveness', which are namely the central get-passives, which are formed by dynamic verbs and have an active counterpart (e.g. *commuters are getting ripped off*), pseudo-get constructions, which seldom have an active counterpart, are not gradable and are either reflexive or reciprocal (e.g. *she didn't really want to get married at that age*), adjectival get-constructions, where *get* is a copula and is followed by an adjective (e.g. *you've got really a bit annoyed at some of it*), idiomatic get-constructions, which are fixed collocations (e.g. *they get accustomed to such kind of environment*) and resultative get constructions, where *get* is followed by a noun phrase before the participle (e.g. *I got my work published in the newspaper*). The data, taken from the spoken part only of the Hong Kong component of the ICLE, show a total number of 151 tokens of *get* followed by a past participle: an interesting first result is that all structures are found in the corpus and also relatively frequently. However, the adjectival is the construction most commonly represented, followed by central passives and pseudo get-constructions, with the idiomatic ones being the least frequent. The same pattern can be seen in the British part of the ICLE, used by the author as a benchmark corpus, while

a comparison with the Indian component of the ICLE, analyzed by the author himself in a previous study, shows some relevant differences, as Indian learners make a remarkable use of central passives.
Influences from the relative L1s of the learners are mentioned by the author as possible explanations, since the passive periphrasis in Hindi is similar to the British one, while it is not present at all in Cantonese.
Moreover, taking into consideration specifically the characteristics attributed to the *get*-passives mentioned above in 3.1, it can be noted how they are present in the instances of this study. In fact, as regards the presence of the agent, it can be found in only three examples in the Hong Kong corpus and in seven and eight in the British and Indian ones, confirming that *get*-passives tend to focus on the patient and are thus mostly produced without an agent; then, as semantics is concerned, it can be seen in the examples of central passives, but also of adjectival constructions, that the connotation is mostly adversative. However, this last result can only be seen in the Hong Kong corpus: this could indicate that some peculiarities of the verb do not apply (or apply less) to some ESL learners.

Due to its many structural patterns and meanings, the verb *get* represents thus quite a challenge for learners. As the general studies in 3.3 and the one in 3.4 have shown, many features of learners' production with high-frequency verbs in general and *get* only appear to be influenced by their specific L1: therefore after having analysed general characteristics of Italian and German learners of English in the second chapter, the next one will be my own research about how these two categories of learners use the verb, in order to see if they present similarities between themselves and with the other learners mentioned in this chapter, and also in order to suggest teaching materials based on the difficulties they seem to show.

CHAPTER 4

4. A STUDY ABOUT THE VERB *GET* ACROSS ITALIAN AND GERMAN LEARNERS

After having analysed some peculiarities of Italian and German learners of English in the second chapter and the main characteristics of *get* in the third one, in this last chapter I provide my own research about how these students produce the verb, in order to see if the previous results about high-frequency, phrasal verbs and *get* itself are in some way confirmed and if some new ones emerge.

4.1 Data and procedure

As seen in the previous chapter, *get* is one of the most important verbs in the English language, since it can form many expressions together with other particles and above all since it can have many meanings, according to the structure and the context in question. This study aims to observe how L1 Italian and German learners of English use this verb, including any phrasal verb or idiom linked to it, in order to detect any over- or under-use, as well as any misuse in the production of written essays. The data are taken from the International Corpus of Learner English (ICLE) mentioned in the second chapter, in particular the second version edited by Granger, Dagneaux, Meunier and Paquot (2009): this one differs from the first 2002 dated version as it is not only restricted to Europe and includes data also from South Africa, China and Japan, with over 1 million more words. The authors define the procedure of data collection as 'rigorous': most of the selected learners are in fact university undergraduates who studied English as a foreign language and the topics regard academic writing and primarily argumentative essays (e.g. "Some people say that in our modern world, dominated by science, technology and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?" or "Discuss the advantages and disadvantages of banning smoking in restaurants"). Moreover, a number of variables that influence their production are

accounted and collected through a questionnaire: these include age, gender, mother tongue background, region, knowledge of other foreign languages, time spent in an English-speaking country, learning context and proficiency level, in order to allow better insights into the differences between sub-corpora and more specific investigations.

Learners' Age and Gender Distribution in ICLE. (Adapted from Granger et al., 2009, pp. 8, 9)

National Corpus (Subcorpora of ICLE)	Average Age	Learners' Gender	
		Percentage Female	Percentage Male
Bulgarian	20,55	83%	17%
Chinese	20,49	64%	36%
Czech	22,07	72%	28%
Dutch	20,75	73%	27%
Finnish	22,73	85%	15%
French	21,70	88%	12%
German	23,39	78%	22%
Italian	24,59	92%	8%
Japanese	20,06	73%	27%
Norwegian	23,94	74%	26%
Polish	23,39	80%	20%

Figure 4: Learners' Age and Gender Distribution in ICLE (Granger et al., 2009, pp. 8-9)

Some observations can in fact be made from the averages of every sub-corpus, as shown in Figure 4: for example, Italians and Germans have a rather high average age (24.59 and 23.39 respectively) and Italians have the highest rate of female students too (92%); besides, the region factor results relevant for those languages that are spoken in more than one country, as for example German which is divided into the Germany, Austria and Switzerland sub-components. Each batch in the corpus represents a number of essays and is identified through a 5-character code, where the first two letters stand for the national sub-corpus, the following one or two stand for the institution code and the last item is the number of the batch (e.g. ITB02 stands for Italian sub-corpus, Università di Bergamo, 2nd batch).

In particular, with reference to the present study, the German sub-corpus (G-ICLE) is made of 437 essays, 229,668 tokens and includes data from the German Universities of Augsburg (GEAU) and Dresden (GEDE), the Swiss one of Basel (GEBA) and the Austrian one of Salzburg (GESA), while the Italian sub-corpus (I-ICLE) is made of 392 essays and 224,222 tokens from the Italian Universities of Bergamo (ITB), Milano – La

Cattolica (ITMC), Milano – La Statale (ITMS), Roma – Libera Università Internazionale degli Studi Sociali (ITRL), Roma – La Sapienza (ITRS), Torino (ITTO) and Università del Piemonte Orientale "A. Avogadro" (ITVE).

Since Italian and German are two non-native varieties of English to be compared, the methodology used in the present study is the Contrastive Interlanguage Analysis (CIA) mentioned in the second chapter. As Granger et al. (2009: 41) point out, "a typical CIA study focuses on one specific linguistic phenomenon – for instance, modals – and has as first stage the automatic extraction of all occurrences of that phenomenon from a native and a learner corpus with the help of a text retrieval software tool": here the occurrences of *get* have been extracted through the concordance program *AntConc*, by typing the search term *get** in the search box of the concordance tool, in order to retrieve the inflected forms of the verb as well (namely *gets* and *getting*), while the search term *got* and its American English equivalent *gotten* were used to retrieve any past form. Successively, the native corpus LOCNESS is analysed through the same procedure, in order to allow a final comparison between the two categories of learners.

4.2 The use of *get* in the I-ICLE

The Italian sub-corpus is a particularly wide one among those of the ICLE, as it is made of 362 essays for a total of 209,524 tokens. However, the number of concordance lines containing the item GET (which includes both forms of *get** and *got*), is not particularly high, as shown in Table 1, where the number of hits is followed by the normalized frequency score (obtained dividing the number of *get* hits by the total number of tokens and then multiplying this by 10,000):

Tokens	209,524
Essays	362
Number of 'GET' hits	198 (9.4 p10kw)
Number of 'GET' + noun hits (possession)	87 (4.6 p10kw)

Number of 'GET' + adjective hits (passive)	48 (2.2 p10kw)
Number of 'GET' + adjective hits (inchoative)	26 (1.2 p10kw)
Number of idiomatic 'GET' hits	26 (1.2 p10kw)

Table 1: figures for the Italian sub-corpus of the ICLE[4]

4.2.1 The possession use of 'get' by Italian learners

With reference to the meanings listed in the previous chapter, it can be observed how *get** in the majority of sentences is followed by a noun phrase and intended in the one of obtaining something (e.g. ITB012002: "*in the end I can say that I am very sceptical about politicians and what they say and I think that they do anything to* ***get people approval*** *again, they always behave in a way that make people have bad opinions about them*", ITTO5006: "*as a result, owning money man had the possibility to lighten his way to build up a decent life going for instance to a good school in order then to* ***get a good job****"*); however, since the same meaning can easily be rendered with the verb *obtain* too, which is more similar to the Italian equivalent *ottenere*, a comparison shows how it is preferred by learners, as it is used slightly more (53 hits of GET, 2.5 p10kw versus 68 hits of OBTAIN, 3.2 p10kw). This sense of the verb is present especially in the crime topic, in recurrent phrases such as 'get money' (e.g. ITRL1006: "*as well as not every person who has serious problems, uses drugs or decides to steal to* ***get money***"), 'get a gun' (e.g. ITRL1006: "*for example, let's try to consider United States' situation: there* ***getting a gun*** *is very easy, but this doesn't mean that a man who owns one must kill somebody*"), 'get a gun license', although only once correctly (0.04 p10kw) and in 13 hits (0.6 p10kw) wrongly spelled as 'licence' (e.g. ITTO2008: "*as the article appeared in the Financial Times shows, in some countries is almost easy to* ***get a gun licence*** *and the deaths by shooting are high*"), or 'get weapons' (e.g. ITTO2030: "*the solution to crime is not create a system that forbid to* ***get weapons****, declaring outlaw who possess them, because, as I said before, people will find a way to have them*").

[4] The number of tokens and essays in Table 1 and 2 is smaller than those mentioned in 4.1 as it refers to the ones which could be decoded by *AntConc*.

However, sometimes the verb in this sense is misused, as another verb would have been correct instead, as *have* in ITTO4005 ("*even when she **gets** babies, a woman should carry on doing all those things that make her happy*"), or it is used in an informal register, as in ITB06003, where the verb *form* would have been a more formal choice ("*if you paraphrase a poem or a novel you **get** a general idea of the meaning of the words, but you can't reach the specific one, since each word has to be considered in the context*"). The past form of the verb *got* is also found to indicate possession through the form *have got* plus a noun phrase in 27 hits (1.2 p10kw), as in ITMC1001 ("*I **have got** really few *real* friends but with all of them I have a different kind of relationship*") or *got* only in 4 hits (0.19 p10kw), as in ITTO5008 ("*ten years ago Jane **got** a first-class honours degree in history at Cambridge University and two years ago she found a job as history teacher in a Secondary School in London*"), with no example produced of the alternative form *gotten*.

4.2.2 The inchoative and passive use of 'get' by Italian learners

After noun phrases, the second most frequent category which is found following *get** is that of adjectives: in the sentences produced by the learners, it takes either the inchoative meaning, as it denotes a change of state (ITTO1002: "*some parents are really unfit for their roles, but most of them are well-aimed mothers and fathers who simply do their best to rear their children. Sometimes they **get confused** by the too many and heavy expectations placed upon them*", ITRS1022: "*the repetition of <*> and so on implies that the two of them are really **getting bored***") or it can form the passive structure (ITTO1015: "*in such a case, even if parents **get asked** to pay for the crimes, the reality would not change*", ITTO4014: "*they would not do it if they thought they were going to **get killed**: they are keen on strong emotions and danger, but they are not suicides!*"). Related to this, in the corpus there are also some phrases which can be attributed to the category of 'pseudo-passives', or 'pseudo get-constructions' as denominated by Coto-Villalibre (2016): *GET married* is the most frequent one with 24 hits (1.1 p10kw), as in ITVE1001 ("*besides, an increasingly number of women are more interested in a successful career than **getting married** and having children*"), then *GET involved*, with 4 hits (0.19 p10kw), as in ITRS2024 ("*countries like Great Britain, Ireland or Denmark, which at first refused of being part of the European communities or like Greece and Spain which had been*

taken out because under a dictatorship asked to ***get involved*** *in it, a deeper awareness of importance of a unified Europe came out*"), with half of them related to the crime topic as well through the phrase 'get involved in(to) crimes' (e.g. ITTO1019: "*they have to cope with life's problems and difficulties, and to realize the reasons why they decided to* ***get involved into crimes***") and *GET divorced* with 1 hit (0.04 p10kw): ITVE1004: "*in this case it is better that their parents* ***get divorced*** *and these children live only with one of them*". Correlated to the family topic is also the phrase *GET pregnant*, which is present 11 times (0.5 p10kw) (e.g. ITTO3028: "***getting*** *artificially* ***pregnant*** *could be felt to be a way for women to to obtain the power of make life decisions*"), which is however twice as much expressed through the more formal alternative *BECOME pregnant* (22 hits, 1.04 p10kw).

In relation to the inchoative use, it can be observable that learners correctly tend to use the verb when they want to express a situation in progress: this can be seen in sentences where *get* is followed by the comparative form of the adjective (e.g. ITRS1038: "*all the characters have their own personality and the author shows their* ***getting older*** *and the changes in the world with the use of symbols*", ITVE2003: "*statistics infact say that the percentage of graduate people is increasing while wellpaid jobs are* ***getting more difficult*** *to find*") or even more intensely by the adverb *more* repeated twice before the adjective (e.g. ITRS2014: "*the temperature of the Earth is increasing every year, 1997 had the highest temperature ever recorded, sun rays are* ***getting more and more dangerous*** *in the hot seasons*"), even though the same concept is expressed more times through the synonym *become* instead of *get* (8 hits, 0.38 p10kw versus 2 hits, 0.09 p10kw). It can also be noted how all instances of progress are not positive at all, since the adjectives that follow have all a negative connotation (*dangerous*, *unbearable*, *difficult*); this is reinforced by no evidence of the phrase *get better*, the comparative form of *good*, while *worse*, the comparative form of *bad*, is present four times in connection with *get**, this as well repeated twice to highlight the deterioration (ITB12004: "*if it doesn't change the state of affairs will* ***get worse and worse****, because after the sceptical attitude, the hard criticism will come out*", ITTO2034: "*it seems to me that the situation is* ***getting worse and worse*** *everyday*").

Moreover, as well as with the possession meaning, with the inchoative one too it can be noted how sometimes another option would have been more pertinent than *get*, as in

ITBO8004, where *come true* would be more suitable than *get true*: "*we can find manager women some rights, some duties, satisfaction for the kind of life they are living and rich in wishes to* ***get true***".

Inchoative and passive structures are found also in the past tense through *got*, as for example respectively in ITTO2041 ("*and then, I* ***got*** *really* ***desperate****, I finally found enough courage to walk into a store, and demanded to see one*") and in ITTO2034 ("**Freeze!* said the man – as we often hear in movies – *Freeze!* he repeated, but the guy moved and he* ***got shot***"), with no example of *gotten* as in the possession use.

The structure *get used to* is also correctly present in their essays, followed both by a noun phrase (e.g. ITB03002: "*people of the thirties or forties generation for instance, who are only now beginning to* ***get used to*** *our frenetic society, consider these last events as a revolution too hard to accept*") and a verb phrase (e.g. ITB02002: "*people have* ***got used to*** *walking less and less because they use car even to go to a place which is five hundred metres distant*").

4.2.3 The movement use of 'get' by Italian learners

The meaning of movement, on the other hand, is present only in one sentence (0.04 p10kw) through the phrasal verb 'get to' (ITRS2007: "*students have to* ***get*** *to University two hours before classes start if they want to find a seat and to listen to the teacher*") and a comparison with *go*, another among the most common polysemous verbs in the English language, shows how learners tend to use the latter to express the same concept (26 hits of *go to,* 1.2 p10kw).

4.2.4 The use of phrasal verbs and idioms with 'get' by Italian learners

The presence of some phrasal verbs and idiomatic forms shows how Italian learners are aware of their various meanings and of how to correctly use them: as for the former, 'get across' is present in its meaning of crossing an obstacle (ITRS1059: "*the simile of the white whale with a wall follows (line 6) and the simile between he himself with a prisoner, either must* ***get across*** *the wall sealing them to be free*"), 'get away' in its main meaning of escaping (ITRL1014: "*this people want to be important and rich always more and once they get involved in a tour of crimes then cannot* ***get away*** *from that*"), 'get away with' in its meaning of managing to avoid being punished for something dishonest

(ITTO2040: "*unfortunately, this events happen because in most of the countries Governments have been **getting away with** the problem of gun ownership for years*"), 'get on' both in its main meaning of having a friendly relationship with someone (ITTO3024: "*it can not be denied that a bad family in which two parents can not **get on** with one another and go on fighting is worst than a family with a single parent in which there is a quiet and relaxing atmosphere*") and in the other of making progress (ITRS2026: "*science is **getting on** day by day in every field, and it is remarkably improving human life*"), 'get out' is present in its main meaning of leaving (ITB10001: "*the sentence shoul be read in its context to be able to **get out** of the labyrinth*"), 'get over' in its sense of dealing with a problem (ITVE1013: "*as far as the particular case of virgins is concerned, it must be claimed that fortunately virginity is no longer a value, it is just a physical factor society must get rid of, because it represents part of the remains of the past male-centered society that we are lethargically **getting over***") and 'get up' in its meaning of leaving the bed after waking (ITB02001: "*if you decide to **get up** at 6:00 in order to arrive at a quarter to seven near your office, you can find a free parking without paying*").

As for idioms, 'get in touch' appears four times (0.19 p10kw), however not with its proper meaning: for example in ITTO1013 ("*young criminals should **get in touch** with their victims, so that they can realize the effective troubles they have caused*") the same concept would have been better expressed by 'have contact with', while in other examples it is wrongly followed by objects, as in ITTO2016 ("*on the contrary, if gun ownership were made illegal, children would have less possibilities to **get in touch** with weapons and become a criminal*"), where 'come into contact with' would have been correct. The idiom 'get rid of*'* appears four times as well (0.19 p10kw) (e.g. ITRS2029: "*despite all this, there are still those who are not able to **get rid of** such a frame of mind which justifies the exploitation of the environment*"), however one instance is incomplete, as the preposition *of* is missing (ITBO3003: "*all over the world, newspapres, televisions, radios have held different views over that cloning experiment and sometimes, in order to sell more copies or to rise television share, the small Dolly has been presented as a little devil of which human beings should **get rid** as soon as possible!*"); 'get to the point' is present once (0.04 p10kw) (ITRS2005: "*the example about poetry is not the only one and maybe is not the most striking **to the point** I want to **get***"), while 'get to know' is present three times (0.14 p10kw), both followed by a noun phrase (ITTO3029: "*for example, one of the major*

*problems for a born *artificially* child is finding his own identity: he could feel a lack of personality or an extreme desire to* ***get to know*** *who is the mysterious donor of the semen*") and by the conjunction *that* (ITTO3011: "*it makes a difference for a child to discover that he or she was born from two people who loved each other, also if afterwards the relationship crumbled, or to* ***get to know*** *that he or she was born in a laboratory and that he or she will never know his or her real father*"). One additional[5] idiom which can be found in the corpus is 'get cross', which is used in the meaning of becoming angry (ITTO1030: "*it is the same feeling a baby would have if its mother allowed it to steal and eat all the candies it desires, without even* ***getting cross*** *with it*").

4.2.5 Other uses of 'get' by Italian learners

Referring to the other meanings listed in chapter 3 and not contained in Table 1, the ingressive meaning is not present at all in the sentences produced by the Italian learners, while the causation one can be found only two times (0.09 p10kw) throughout the corpus, through the structure get + noun + past participle (ITTO4015: "*frequently in fact they refuse to reflect about what would be of their children if they* ***got themselves killed*** *or seriously injured*") and get + noun + to-infinitive (ITB06004: "*another positive aspect of explaining a text and of reading the critics' explanations is that the reader can compare his point of view, which is a way of* ***getting new ideas to develop***"). However, since this meaning can be found in many more instances with other verbs such as *let* (e.g. ITRS1069: "*this is a very interesting technique, which let the poet show something instead if saying it*") or *make* (e.g. ITTO1021: "*whatever the case, the proposal of making families paying compensations to the victims of their children's crimes, sounds like a curious way to attract parents' attention to the problems their sons have*"), it can be inferred that learners do have knowledge of the structure itself, but need more practice in its use through *get* in particular.

Moreover, the obligation meaning of the verb expressed through the past form *got* is also known by learners, as the four hits (0.19 p10kw) of the form *have got* plus *to-infinitive* show (e.g. ITMC1001: "*in the real world every person has got his own way of thinking, their own perception of reality and even friends* ***have got to deal*** *with different*

[5] "Additional" referred to idioms means that they are not listed in the third chapter.

personalities which sometimes choke even in the closest of the relationships", ITRL1017: "I think, first of all, that we ***have got to distinguish*** *the phenomenon *Violence in the stadiums* in its display in the British countries from the manner it show in latin ones"*); however, this form is a variation of the more common *have to* only plus *to-infinitive*, which is found in fact relevantly more frequently throughout the corpus (183 hits, 8.7 p10kw), as for example in ITTO2024 ("*first of all, we have to consider how the problem of gun ownership goes beyond the phenomenon of organized crime*").

4.2.6 Errors produced by Italian learners

Although the students produced instances of all the senses mentioned above of *get*, it needs to be pointed out how some of their texts present mistakes: sometimes they regard a L1 interference, as in ITTO2040 ("*lastly, the use of arms should be forbidden because people who suffer from depression and* ***get an arm*** *could be more tempted to commit suicide than others people*"), where 'get a weapon' is wrongly expressed by 'get an arm' because of the Italian translation of 'weapon' *arma*, while others regard on the other hand grammatical structures. For example, taking into consideration the structure *get used to*, there are two examples (0.09 p10kw) where it is produced followed by the verb in its infinitive, and not the -ing, form (ITTO5007: "*when someone* ***gets used to buy*** *everything this becomes a sort of *illness**", ITRL1019: "*while the people living in the Eastern countries have always* ***got used to make*** *sacrifices for the State suppressing their own individuality, the Americans have always privileged it*"); *worse*, the comparative form of *bad*, is also subject to a number of mistakes, as more than once is confused with its superlative form *worst* (ITRL1014: "*the society is* ***getting worst*** *than before, and guns, drugs, alchool, robberies are rasing up*", ITTO5007: "*someone tries to improve its state with its own strenghts, other* ***get worst*** *and making money becomes the only reason of their life*"). Furthermore, in one instance (0.04 p10kw) the phrase *get divorced* is wrongly expressed, as the verb following *got* is in its present and not in its past form, probably meant as a noun phrase but with the article missing (ITTO3006: "*there are lot of evidence of women and men carrying one or more children from a relationship that is not the *holy marriage*, and a lot of children from one family can meet the children from another family because their parents* ***got divorce****, fell in love with a new person and go to live together with the *new* family*").

Some other sentences do not sound right for how they are structured: for example, in ITTO6007: "*on recent times the number of couples who decide to divorce is highly increasing, alongside with the number of couples who choose to cohabit without marrying and of people who* ***got*** *for living as singles even though with children to bring up*", *get* is redundant, as the sentence sounds correct with the verb *live* alone, in its present form ('people who live as singles…').

Phrasal verbs and idioms are sometimes used in a wrong way too: in addition to the misuses of 'get in touch' and 'get rid of' mentioned above, the phrasal verb 'get back', which is present once (0.04 p10kw), is used in a way that does not sound English and makes the sentence wrong (ITRS2030: "*fortunately good sense and the fear of a war* ***got*** *them* ***back*** *such intentions and everybody could take breath*"), as the idiom 'get in(to) trouble', which is present one time in the corpus as well (0.04 p10kw), although it turns out wrong as the noun is expressed in its plural form (ITTO1006: "*children living in a complicated society like ours, where it is even a problem to go out alone because of the risks that it implies, should have always the possibility of learning from other important social figures such as pedagogists or sociologists what is worth to do and what is not, what is good and what is bad for them, how to satisfy curiosity without* ***getting in troubles*** *and how to channel one's aggressivity due to familiar contrasts*").

4.3 The use of *get* in the G-ICLE

The German sub-corpus too is among the biggest of the ICLE, as it contains 378 essays for a total of 197,681 tokens; this comes with a high number of GET hits as well, as can be seen in Table 2:

Tokens	197,681
Essays	378
Number of 'GET' hits	597 (30.2 p10kw)
Number of 'GET' + noun hits (possession)	188 (9.5 p10kw)

Number of 'GET' + adjective hits (passive)	49 (2.4 p10kw)
Number of 'GET' + adjective hits (inchoative)	98 (4.9 p10kw)
Number of idiomatic 'GET' hits	168 (8.5 p10kw)

Table 2: figures for the German sub-corpus of the ICLE[6]

4.3.1 The possession use of 'get' by German learners

In the majority of the instances of the corpus, *get* is followed by a noun phrase and takes the meaning of possession (e.g. GESA5005: "*we head there, speak to the manager and **get a description** of the robber and the goods he stole*"); the phrases formed with this structure that are more recurrent are 'get the/an impression', which appears 14 times (0.7 p10kw) (e.g. GESA3004: "*I constantly **get the impression** that many people live their lives only in one direction*", GEBA1004: "*they help students to **get an impression** of the past and the present and allow them to see the world through somebody else's eyes*"), 'get the chance', which appears 8 times (0.4 p10kw) (e.g. GEDR1013: "*some can resist and will not take the wallet that does not belong to them, but others find it quite thrilling, and would take it and try it again if they **get the chance***"), 'get information', which appears 5 times (0.25 p10kw) (e.g. GEBA1028: "*although TV is not at all the only means to **get information**, amusement and maybe even the answers to life's problems, it has become a very influential medium during the last few decade*") and 'get money', which appears 4 times (0.2 p10kw) (e.g. GEBA1048: "*music is not only used to **get money**, to present one's voice and body, but there is still a "pure" side of music that has survived until today*"). Furthermore, it can be found often in the school and work topic, in phrases as 'get a job', which is present 6 times (0.3 p10kw) (e.g. GESA3013: "*in former days girls did not often have the chance to enjoy education or to **get a job** beside traditional professions like chambermaids or housemaids*") or 'get an education', which is present 4 times (0.2 p10kw) (e.g. GEDR1012: "*rich people, for example, can afford to send their*

[6] The number of tokens and essays in Table 1 and 2 is smaller than those mentioned in 4.1 as it refers to the ones which could be decoded by *AntConc*.

*children to privat schools where they can **get a high education**, and will therefore have the highest chances on the job market*").

In this sub-corpus as well, the past form *got* is used to express possession through *have got* plus a noun phrase (38 hits, 1.9 p10kw), as for example in GEAU1044 ("*my aunt **has got** a business of her own, which means that my young cousin is given 5 Marks a day to buy some junk food at McDonald's*"); this form is also found, even if less frequently (26 hits, 1.3 p10kw), without the verb *have* (e.g. GEAU3099: "*I very nearly **got** a chance to learn how to use a fire extinguisher the day when my Chinese flatmate tried to make jacket potatoes*"), with no example of *gotten*.

4.3.2 The inchoative and passive use of 'get' by German learners

As well as nouns, *get* is also followed by adjectives in many instances: the structure can have an inchoative meaning, when it denotes a change of state (e.g. GESA3004: "*I have noticed that many pupils **get aware** of their 'future-orientation' for the first time when they finish school*") or it can form a passive (e.g. GEAU1064: "*in Europe though there are many self-confident, independent women. They want to behave like men but at the same time get complements, have doors held open for them and **get served** first*"), and the two uses are found with *got* as well, as respectively in GEAU3058 ("*I started and was very enthusiastic about cycling but very soon I **got tired** and began to sweat*") and in GESA4008 ("*in the second world war it was the Jews who **got killed** in the concentration camps*"). The category of 'pseudo-passives' is present in the German sub-corpus too, as the phrase 'GET married' is found in 16 hits (0.8 p10kw) (e.g. GESA5023: "*after five weeks both were allowed to leave the hospital and as they had learned to love each other, they decided to **get married***"), 'GET divorced' in three hits (0.15 p10kw) (e.g. GEBA1027: "*nowadays if a couple decides to get married the odds are two to one that they will end up **getting divorced***" and 'GET engaged' in one (0.05 p10kw) (GEAU1001: "*to let the cat out of the bag, we plan to **get engaged** this year in summer and I can well imagine being her husband one day*". In relation to the inchoative use, 'GET pregnant' is also present in 2 hits (0.1 p10kw) (e.g. GEAU3016: "*for my part I solved the problem, because I **got pregnant** and therefore I shouldn't smoke too much*"), as well as the use of *get* for expressing an action in progress, which can be found in the production of the verb followed by a comparative form of the adjective (e.g. GEAU3003: "*dare devils, jay-*

walkers and rebels are fine when they are still sweet slumbering babies, or firmly on a leash when they ***get older***").

With reference to this, it can be noted how German learners also tend to repeat the comparative form twice in order to better express the change of state through the verb (e.g. GESA5006: "*after dinner at 18.30 I usually read a book. It's* ***getting harder and harder*** *to find one I don't know yet*" and, in particular, *get* followed by *more and more* plus an adjective can be found in 8 hits (0.4 p10kw) (e.g. GEAU1078: "*the Somalians seem to* ***get more and more fed up*** *with the UN, even the peaceful people, the atmosphere is changing from salvator to occupier*") and not only followed by adjectives with a negative connotation, as showed in GEBA1027 ("*I found myself being rather bored after a short time while he was* ***getting more and more excited*** *and started shouting and jumping up and down hoping to see the horse he had bet ten punts on come first*"). Moreover, the verb is also found where two comparative forms are compared (e.g. GESA5004: "*the later it* ***gets*** *the more often John looks at the watch*") and a tendency to use it when expressing negative situations can be seen in them as well, as they use *get* together with the comparative form of *bad*, *worse*, in three hits (0.15 p10kw) (e.g. GESA5018: "*now I know that we are forced to do whatever they want us to do, otherwise the situation is* ***getting worse***"), with no instance of the verb followed by *better*, the comparative form of *good*.

Finally, the structure *GET used to* can be found in 15 hits (0.7 p10kw), both followed by a noun (e.g. GEAU4011: "*this only started to improve after three or four weeks when I had* ***got used to*** *the peculiarities of different accents and dialects*") and a verb (e.g. GEAU3075: "*people, especially male ones,* ***get used to*** *seeing a point of honour in having the ultimate form of wisdom about things like right angles and 0.01% more profitable investments*").

4.3.3 The movement use of 'get' by German learners

GET is used for expressing movement in 51 hits (2.5 p10kw) throughout the corpus: this meaning can be conveyed through some phrasal verbs (which will be analysed in detail in the next section), namely 'get to' (e.g. GEAU2009: "*on a arctic Saturday afternoon in December, when you shovel your way through crowds of stale smelling coats and dirty jackets, you try to* ***get to*** *your warm and cosy room as quickly as possible by walking in*

the street"); 'get into' (e.g. GEAU3001: "*Steffi, however, comes home from school, needs her key to* ***get into*** *the flat – her parents are both employed and she's the only child*"); 'get out' (e.g. GEAU3082: "*anyway, to compensate for the boredom of Reading, I grabbed every opportunity to* ***get out*** *of the place*"), especially in the
expression 'get* out of bed' which is found three times (0.15 p10kw) (e.g. GESA5004: "*he* ***gets out of the bed*** *with verve*"); 'get on' (e.g. GEAU3082: "*hurriedly I changed my ticket and hopped on the train for Victoria Station, which meant that I had to change there and* ***get on*** *the Underground, which was supposed to bring me out to Knightsbridge*"); 'get off' (e.g. GEAU1099: "*Have you ever suffered the pain of home-cooking? Dragging large, heavy weighing bags from Sainsburys to the bus stop, being overseen by a reckless cyclist when* ***getting off****, with all the food – eggs, tomatoes, tuna steaks and honey – scattered all over the place?*") and 'get away' (e.g. GEAU3060: "*and one of them is bound to take the seat next to yours – no chance of* ***getting away***"). Among the other expressions to express movement, those that recur more often are 'get home' (e.g. GEAU1079: "*when she* ***gets home****, the whole family is already awaiting her*") and 'get there' (e.g. GEAU2031: "*did you ever forget to buy fresh bread and the best bakery in town is kilometers far away? How comfortable is it to use a bus to* ***get there****!*"), which both appear three times (0.15 p10kw).

4.3.4 The use of phrasal verbs and idioms with 'get' by German learners

The German sub-corpus presents a high number and variety both of phrasal verbs and idioms, a fact that seems to indicate how they represent an area of little difficulty for learners. As for phrasal verbs, the ones which can be found in the corpus are: 'get ahead', in its sense of progressing (GESA5045: "*obviously one can combine possibility one and possibility two so that one* ***gets ahead*** *more qoickly and to obtain better results*"); 'get along', in its sense of having a friendly relationship, both with people (e.g. GESA5036: "*people who know what they want and care about how others expect them to behave try to* ***get along*** *with others",* with abstract things (GEDR1013: "*however there are also other people who commit crimes, because they do not* ***get along*** *with their social circumstances*" and with no object following the verb as well (e.g. GEDR1021: "*the technically equipt life offers many opportunities to* ***get along*** *without the help of others*"; 'get away', in its main meaning of escaping (e.g. GEAU1079: "*they were trying to* ***get away*** *from stress*

and smog, they longed for peace and quiet in the suburbs or even a bit further out and headed towards the blossoming orchards of Kent or the fertile Southcoast regions near Portsmouth or Brighton", GEAU3084: "*although kids don't intend to upset their parents they use every opportunity to* ***get away***"); 'get back', in its senses of returning where everything started (GEAU3035: "*I went through a process of changing my style every year and, I think, finally* ***got back*** *to my very own way of writing with which I had started secondary school*") and of getting something that was lost (e.g. GEAU3049: "*a friend of mine first went to the body building center to* ***get*** *his muscles* ***back****, after he had broken his leg*"); 'get into', in its sense of being involved in something (GESA4011: "*it has been proven, that he had been treated unfairly, but the American Justice does not want to* ***get into*** *the case once again*"); 'get off', in its senses of dismounting from a bike (e.g. GEAU3041: "*all you can do:* ***get off*** *your comfortable bike, push your vehicle and join the mass of pedestrians*") and of getting out of a train (GEAU4013: "*another question that has always bothered me is why do old people have to leave their seats two stops before they finally* ***get off****?*"); 'get on', in its senses of having a friendly relationship with someone (GESA5021: "*to* ***get on*** *well with others, to give love, to take sometimes a couple of minutes only to look and to listen, far away from everyday' s noise*"), of boarding a train, or a plane as well (e.g. GEAU3082: "*from there I* ***got on*** *the Underground out to Munich Airport, missed the last bus shuttle for terminal 2 and therefore nearly missed the plane,* ***got on*** *the plane, landed in Heathrow airport about one hour later*"), or of making progress (GEAU3047: "*before entering into an agreement he also goes to see his astrologer to be told how to proceed. An extraordinary way to* ***get on*** *to the top, isn't it?*"); 'get out', in its senses of leaving (e.g. GEAU2012: "*people don't notice each other, simply hurrying from one place to another in order to* ***get out*** *of here as quickly as possible*"), escaping as referred to prisoners (GESA4011: "*there are also people imprisoned and wait for their execution who are innocent. They never had a fair lawsuit and a real chance to* ***get out*** *of it*") and of spending time outside home (e.g. GESA5021: "*a sunbeam tickles your nose. It's like a hint of the sun, meaning: "Come out of your house, walk around,* ***get out*** *in the fresh air!"*"); 'get over', in its sense of recovering (e.g. GEAU3047: "*sitting in the back row of the bus on their way home from school, they hold their heads parting to parting over the page where Libra, Sagittarius and Aquarius are informed about their destinies of the following week: "Love: You will* ***get over*** *a short lover's grief... "*"); 'get

through', in its sense of contacting someone by telephone (GEAU3031: "*"By now at least the telephone works better," he told me recently, "in the beginning I sometimes had to try nearly all day until I could* ***get through****'"*"); 'get to', in its sense of reaching a place (e.g. GEAU1017: "*of course, a young man from a small village in the country says he couldn't afford not to have a car of his own because he has to* ***get to*** *the city for everyday's work*"); 'get together', in its sense of meeting (GEAU3010: "*I was usually terribly bored when they* ***got together*** *talking about nothing else but shower units and wall-to-wall-carpeting*") and finally 'get up' in its sense of leaving the bed (e.g. GEAU3058: "*the only reason to* ***get up*** *in the morning at the moment is the fact that I will go to the beergarden later on*"). As for idioms, some of the most common ones can be found a relevant amount of times, as well as some others, although only once. These are in particular: 'get a picture' (GEAU2036: "*so let's look at some advantages and disadvantages to* ***get a picture*** *of this problem*"); 'get carried away' (GEAU1067: "*the wonderful Greek restaurants where so many people* ***get carried away*** *memorizing their wonderful holiday in Crete*"); 'get in(to) contact', which is present four times (0.2 p10kw) (e.g. GEBA1048: "*my friend, interested in native people and their rights, saw his chance to* ***get in contact*** *with them*"); 'get in touch' (GEBA1036: "*the learner* ***gets in touch*** *with another culture, other people and another world, that is, he enlarges his horizon*"); 'get into trouble', which is present two times (0.1 p10kw) (e.g. GEDR1022: "*as a child you have no real responsibility, you actually can try out whatever you want to try and there are enough people taking care for you, such as parents who will definitely help you if you should* ***get into trouble***"); 'get on someone's nerves', which is present four times (0.2 p10kw) (e.g. GEAU1023: "*it has been* ***getting on my nerves*** *that my floor mates could listen to every single word I was saying*"); 'get rid of', which is present 19 times (0.9 p10kw) (e.g. GEAU1043: "*sitting behing the wheel of a fast car makes them feel strong and superior and they try to* ***get rid of*** *their aggressions with speeding and illegal racing*"), although in one example is wrongly expressed by 'get rid off' (GEAU3008: "*it became clear to me that the only chance to* ***get rid off*** *her guidance was to prove that I was old enough to look after myself*"); 'get there' (GEBA1017: "*the results are sometimes rather sad: those women become very hard and are often more ruthless than men would be in a similar position. But then this is not surprising with the struggle of* ***getting there*** *being so much harder*") and 'GET to know', which is present 27 times (1.3 p10kw) (e.g. GEAU1075: "*being*

allowed to visit a course with about thirty people you surely will ***get to know*** *new fellow students who will help you with your problems*").

4.3.5 Other uses of 'get' by German learners

In the German sub-corpus both the ingressive and the causation uses are present: as for the former, it can be found in four hits (0.2 p10kw) (e.g. GE-AUG-0051.1: "*true, chips with vinegar are a tonguesore, but try them with mayonnaise as some of the English do, you will* ***get to like*** *them*") and as for the latter, it can be found in 25 hits (1.2 p10kw), with most of the structures through which it can be expressed (e.g. GE-DRE-0001.1: "*a clue to that problem is to* ***get pupils motivated*** *for doing their homework*"; GE-AUG-0086.3: "*if, contrary to the rule, certain children do not get lost when running free, they are at least sure to* ***get themselves and their parents into trouble***"; GE-AUG-0014.3: "*as long as we don't accept that getting involved inevitably means* ***getting our hands dirty****, starvation will not come to an end in the Third World*", GEAU1084: "*even if you use the break to go to the fridge to* ***get yourself a cool drink*** *and hopefully no sweets, the very special atmosphere of this moment which means shiny eyes and handkerchief is lost forever*").

Moreover, the modal meaning of the verb through the form *have got* plus *to-infinitive* is known by German learners too, as it appears 9 times (0.45 p10kw) (e.g. GEAU3097: "*I got up at 6 o'clock this morning because Phil told me that you'****ve got to register*** *for the practical in the lab after your third semester on the very first day of your first semester if you want to get a place*").

4.3.6 Errors produced by German learners

With reference to the structure *get used to*, in just one sentence (0.05 p10kw) it is produced wrongly with the verb in its infinitive form instead of the -ing one (GEBA1053: "*one should think Phil I students* ***get used to*** *write essays, that they perform a ritual like thing and just write down their 500 words. But unfortunatelly to write an essay is no program one could click on in his / her brain*").

One error regarding on the other hand a German interference can be seen in GEAU3016: "*when I recognised that I'm* ***getting a baby****, I still woke up every night, but what I did and what I'm still doing is that I went to the toilet quickly and returned to bed, closing my eyes*

again and sleeping immediately", where the concept of being pregnant, which could have been rendered by "expect" or "have a baby", is wrongly translated from the German "ein Kind bekommen", where *get* is the main equivalent of the verb *bekommen*. There are other instances where the verb *get* is used inappropriately, as another verb would sound more natural in the English language: this is the case of GEAU3049: "*he really **got** a different person, the only thing he ever wanted to talk about was his body and the progress he made*", "*but within two or three months he really **got** a "bodybuildaholic". He went there every evening to strain his muscles, to work at silly machines, following a strict plan*" and GESA5029: "*taking meal or dinner is **getting** more and more just a method to appease hunger-pangs*", where respectively *became* and *becoming* would have expressed the concept in the right way. This could derive from the learners' wrong assumption that *get* is a synonym of *become* not only when followed by adjectives, but also when followed by nouns; however, this does not seem a recurrent mistake, since the examples come from the production of only two learners. One other instance is in GESA5014, where *have got* is wrongly used instead of the verb *have* only: "*during dinner at about 7.00 p.m. he asked me what day we **had got***".

Another error regards the use of the comparative, in particular the minority one, as the adjective is not expressed by its bare form ('funny') after *less* (GEAU2040: "*nosy neighbours can be a reason for life **getting less funnier***").

Finally, the idiom 'get to know', in its many instances present in the sub-corpus, includes some wrong structures, showing how learners do not have a clear image of how it is correctly used in the English language: this can be seen in GEAU1075: "*but it is not only the new students who **get you to know** it is also the professor who often **gets to know** his students*", where the second instance is right, while the first is not, since the object is put in the middle and not after the idiom ('but it is not only the new students who get to know you…'), and in GESA5023: "*after one week they **got known** each other much better and they realized that it had been bad to live too healthy as well ass to eat too much without doing some sport*", where not only *to* is missing, but also *know* is wrongly expressed in its past form, while to correctly convey the past tense of the sentence just *get* has to be in its past form ('after one week they got to know each other…').

4.4 A comparison between the Italian and German learners

	Italian sub-corpus	German sub-corpus
Tokens	209,524	197,681
Number of 'get' hits	198 (9.4 p10kw)	597 (30.2 p10kw)
Number of 'GET' + noun hits (possession)	87 (4.1 p10kw)	188 (9.5 p10kw)
Number of 'GET' + adjective hits (passive)	48 (2.2 p10kw)	49 (2.4 p10kw)
Number of 'GET' + adjective hits (inchoative)	26 (1.2 p10kw)	98 (4.9 p10kw)
Number of idiomatic 'GET' hits	26 (1.2 p10kw)	168 (8.5 p10kw)

Table 3: figures for the two ICLE sub-corpora compared

Table 3 shows a comparison between the ICLE sub-corpora in terms of number of hits and of normalized frequency scores, which provide more detailed figures on which the first observations can be made. What stands out the most is in fact that even though the total number of words of the two does not differ much, there is a big difference regarding the number of times *get* is found in the essays of Italian and German learners, with the latter producing the verb more than three times as much. This leads also to a wider variety of uses of the verb by German learners: they present in fact a relevant number of instances of all the uses mentioned in Chapter 3, while Italian ones show either a lack of production (for the ingressive meaning), or just one instance of others (movement and causation), which, being both produced by an individual learner, cannot be considered of common use by the whole category.

The most significant difference between the two regards the quantity and variety of phrasal verbs and idioms with *get*, since German learners use and therefore seem to know many more of them and, in reference to the phrasal verbs only, often more than just the

main meaning. In particular, the idiom *get to know* is highly frequent in the corpus, as it appears in a number of hits higher than the total number of phrasal verbs and idioms in the Italian sub-corpus (27 hits, 1.3 p10kw versus 26 hits, 1.2 p10kw): this could derive from the German expression *kennen zu lernen*, which conveys the same meaning and is frequently used in the German language as well. Similarities on the other hand can be seen when considering the possession meaning, since it is the most frequent in both categories, and the passive structures: in fact, even though the wide difference in terms of the total hits of *get* previously mentioned, the two categories of learners produced the same amount of them.

As Lorenz (2009) explains, in Germany all pupils are required to learn at least one foreign language: Russian used to be the predominant one and then during the former division of the country, English gained more importance, initially only in the West and, after some difficulties in terms of traveling in the days of the GDR, also in the East; over the last decade English has become the main one taught, since "in 1994, 96% of all pupils had learnt English at some stage in their schooling" (Lorenz 2009: 132).

As regards university, the author describes how students, in order to enter, need to take the so-called *Abitur*, which corresponds to the British A-levels: however, German students need to take many more subjects in addition to their two preferred ones and since 1997 one foreign language, which is usually English, has to be taken, meaning a total of five hours per week. This educational system, which is also often followed by an encouraged period studying or working in an English-speaking country, can thus show the importance given to the English language and can be a key factor for the abundance of *get* uses found in this study.

On the other hand, Italy has a different situation, as Prat Zagrebelsky (2009: 139) explains: being under the Fascist regime until 1945, the country went through a linguistic policy which banned the use of foreign words and still in these years some Italian people think that the use of any English term is a sign of cultural subordination to the USA; however, since the 1980s, English has become the most popular foreign language studied at school and is taught in both primary and secondary school.

Taking the university environment into consideration, the fact that the degree in Foreign Languages and Literature tends to be more literature oriented, with less relevance to the practical use of the language and with no compulsory attendance, could be a factor which

explains the difference between Italian and German learners in their use of the language. However, a recent additional degree course in English Language and Linguistics and a rising percentage of students doing exchange programmes abroad seems to indicate a change, as Italian students will possibly have more exposure to the language.

As for errors, no major differences can be pointed out: both learners present in fact some due to the interference with their mother-tongue and both seem to have the same difficulties with the difference of use of the verb when followed by the infinitive and by the *-ing* form; moreover, they both present some syntactical mistakes, especially in the structure of comparative forms.

In the light of these problems, both categories of learners, in relation to the massive topic of high-frequency verbs, could take advantage of the use of corpora in their classroom already in their first years of university: starting from their own mistakes, which could be collected by the teachers from the essays produced during their exams, they could then analyse them together and make a comparison with the correct versions in native corpora, in order to practice with these structures and avoid to think in their native language from the beginning.

With reference to Italian learners only, they show a smaller variety of uses, especially of phrasal verbs and idioms, which seems to confirm the findings of Prat Zagrebelsky (2004) mentioned in the second chapter; since these structures are also often misused, they could focus especially on these when examining examples from native corpora; a supplementary task in the classroom could involve the use of dictionaries, where learners first study the main entries with examples and then are asked to provide sentences autonomously for each meaning they remember, in order to broaden their knowledge in terms of meanings until they seem to master the most important ones. Additional exercises for practice could be of the 'fill in the blanks' type, where only the verb is provided and students have to add the particles which form the rest of the phrasal verb and vice versa, so that, by seeing the different contexts of each example, they can learn and remember more and see how to produce them. A final part where they provide examples of their own needs also to be added, mostly because these verbs tend to be produced mostly orally and it is thus important to know the situations in which they can be used. Starting from the most common high-frequency verbs such as *get* and the related

phrasal verbs and idioms seems a good option, since as can be seen advanced learners, who as mentioned above have a rather high average age too, still present difficulties.

4.5 The use of 'get' in the LOCNESS native corpus

The last part of this study about *get* involves the analysis of the native corpus LOCNESS: this comprehends two sub-corpora of essays written by British and American university students, which are both argumentative and literary. This allows therefore to see how students of English L1 use the verb in more contexts and also to make comparisons with the learner sub-corpora above.

4.5.1 The use of 'get' by British students

Tokens	155,891
Essays	204
Number of 'GET' hits	112 (7.1 p10kw)
Number of 'GET' + noun hits (possession)	35 (2.2 p10kw)
Number of 'GET' + adjective hits (passive)	16 (1.02 p10kw)
Number of 'GET' + adjective hits (inchoative)	8 (0.5 p10kw)
Number of idiomatic 'GET' hits	32 (2.05 p10kw)

Table 4: figures for the British sub-corpus of the LOCNESS

As can be seen from Table 4, the total number of GET hits is not particularly high in the British sub-corpus, however a variety of meanings can still be analysed. It is possible to find the movement one, especially in the argumentative essays related to transport (e.g. Transport 02: "*to **get from A to B**, they stop at, and sometimes divert several miles to reach C, D and E*", Transport 13: "*many people also feel unsafe to walk or to use public*

transport to ***get to places***") and through the phrase 'get to work', which is present 3 times (0.19 p10kw) (e.g. Transport 03: "*for a minimum of two hours everyday the roads surrounding big cities are packed solid with people trying to* ***get to*** *and from* ***work*** *or school*").

When followed by a noun phrase, *get* is used in its possession meaning, which can be expressed both passively (e.g. Boxing – B03: "*this discussion arose after lots of boxers started to* ***get serious injuries*** *and brain damages as a result of boxing*") and actively (e.g. ICLE-ALEV-0027.8: "*many millions will be made by the genetic engineering of crops and yet the scientists would never* ***get any money*** *and very little credit for it*").

On the other hand, when followed by an adjective, it can form the passive, which is present especially in the argumentative essays about fox hunting (e.g. Fox hunting – FH03: "*however it is still perfectly legal for a group of people to chase a fox with a pack of dogs and watch them chase the fox until it is exhausted and then* ***get torn apart*** *by the dogs*") and boxing (e.g. Boxing – B08: "*boxers, and boxing fans alike, argue that an average fight may only last six 3 minute rounds, and during this they may only* ***get hit*** *100 times in this 18 minute period*"), or it can denote an inchoative meaning (e.g. ICLE-ALEV-0001.7: "*when the press* ***got bored*** *of this story, the bug dissappeared and with it peoples worries about its effects*"), especially through the comparative form of the adjective (e.g. ICLE-ALEV-0004.7: "*should the attitudes of the people involved in this persevere or* ***get stronger*** *however, then there could be some major problems ahead*").

However, in this sub-corpus *get* is found with a higher frequency in its causation meaning and through phrasal verbs and idioms: the former can be seen in 10 hits (0.6 p10kw) through different structures, such as *get* + noun + to infinitive (e.g. ICLE-ALEV-0012.9: "*the Ministry of Agriculture would be expected to clarify once and for all whether it was safe or not and then prove it to try and* ***get people to eat beef*** *again before it became too late*") or *get* + noun + adjective (e.g. ICLE-BR-SUR-0028.1: "*it was his job to* ***"get his hands dirty"*** *in the words of Sartre and actually implement the policies evolving from the President*").

As for phrasal verbs, those produced by the learners are: 'get across', in its meaning of making other people understand a message (ICLE-BR-SUR-0001.1: "*Camus often used tragedy to get across his ideas*"); 'get around', in its meaning of moving around (Transport 03: "*travel has become and essential part of our everyday lives, we use it to*

go to work or school and just to ***get around*** *in general because it's easier and quicker than walking*") and of avoiding a problem, expressed by the more British alternative 'get round' (Parliamentary system 01: "*the main way seen in* ***getting round*** *this problem is by adopting a single trans double vote system*"); 'get back', in its meaning of returning (ICLE-BR-SUR-0006.1: "*when he* ***gets back*** *however, he finds that he likes life much and refuses to return to death*") and of returning someone somewhere (Transport 09: "***getting their employers back*** *to work will lead to an increasing standard of railway journey and stability of trains*"); 'get back at', in its meaning of gaining revenge (ICLE-BR-SUR-0001.1: "*joining the party seems to have been an act to* ***get back at*** *his father, perhaps humiliate him, rather than one which was seriously thought out and carried out from conviction belief in the party's line*"); 'get into', in its meaning of being involved (006: "*yet these people would* ***get into*** *a vicious circle where the only way to pay off gambling debts would be to win, and so the problem increases*") but also of being accepted to study at a university (ICLE-BR-SUR-0016.1: "*to* ***get into*** *one of these highly prestigious colleges you have to take another exam as well as having the bac, rather like oxford and Cambridge*"); 'get off', in the sense of moving (something) away (008: "*also many sporting projects receive grants to build more recreational areas for the youth and public to express themselves and* ***get*** *them* ***off*** *the street and decipline them*"; 'get on', in its sense of having a friendly relationship (e.g. ICLE-BR-SUR-0032.3: "*simplicity is one of the Commission's key aims, with the idea that if things are simpler and more easy for everyone to understand people will be happier,* ***get on*** *better*"); 'get out', in its sense of leaving (e.g. *when it comes down to it an individual is an individual with his own mind and is able to* ***get out*** *of a contract if he so choses*"); 'get through', in its meaning of making someone understand something (Transport 11: "*a huge advertising campaign explaining how much cheaper rail travel is when compared to other modes of transport would eventually* ***get through***"); 'get to', in its sense of reaching a place (e.g. ICLE-BR-SUR-0024.2: "*in an attempt to* ***get to*** *Venice, Candide loses his last sheep in being cheated*"), which can be meant also metaphorically (e.g. ICLE-BR-SUR-0007.1: "*his quest has been to discover his project and he undergoes certain tests to* ***get to*** *his position of authenticity and liberty at the end*") and 'get up' in its sense of leaving the bed (e.g. ICLE-ALEV-0008.8: "*having a child is stressful at any age but if the mother is older when the child very young she may not have the energy to* ***get up*** *four times every night*".

As for idioms, three can be found in the sub-corpus: ‘get on with life’ (ICLE-BR-SUR-0017.2: “*he becomes pragmatic and realizes that man should **get on with life***”), ‘get out of hand’ (ICLE-ALEV-0004.7: “*the politically correct "loony" lobby seem to have got their weight behind this movement and when this is coupled to the hysterical reaction over things such as the veal-crate??? fiasco last year, things **get out of hand***”) and ‘get rid of’, which can be found three times (0.19 p10kw) (e.g. ICLE-ALEV-0012.9: “*farmers would have to **get rid of** their stock at great expense and try and farm something else which would be suited to their land and area*”).

As for the past form of the verb *got*, it can be highlighted how it is not a recurrent item in the sub-corpus, since it is present only in 10 examples (0.6 p10kw) throughout the whole of it: seven of these (0.4 p10kw) express possession, through the form *have got* + noun phrase (e.g. ICLE-BR-SUR-0024.1: “*the text is more adhered to now as Mitterrand **has not got a majority** to back up his policies in Parliament but he still enjoys the privileges and wields power in the same manner as his predecessors*”), with two among them (0.12 p10kw) expressed by *got* only (e.g. ICLE-BR-SUR-0008.1: “*in the court he often felt superior to the judge and he enjoyed **the satisfaction he got** from looking down on people and judging theme*”), while as regards the modal form have got to + infinitive, it is present in just one instance (0.06 p10kw) (Transport 12: “*to improove the rail service trains **have got to be timed** to arrive and depart at key times i.e. arrive at eight o'clock and leave at half past six*”), with no instance of the form *gotten*.

One last use of the verb which can be found two times (0.12 p10kw) is the ingressive (e.g. Fox hunting – FH01: “*the riders rarely actually **get to see** the fox, once it has been caught*”).

In this sub-corpus as well are present some examples where an alternative for *get* would be better, even if learners are of English L1: for instance, in Transport 02, the verb *take* would be a better option than *get* (“*closing small lines which regularly lose money may seem sensible but leads to less people **getting** trains in general*”) or in Boxing – B05, *move* instead of *get away* would sound better (“***getting away** from the long term effects to the short term effects, yes brain cells are still being killed but when a boxer receives a very strong punch direct to the face not only would it damage his nose, eyes, teeth and general facial features but it could also cause him to enter a coma, haemorage or die*”).

4.5.2 The use of 'get' by American students

Tokens	178,109
Essays	232
Number of 'GET' hits	305 (17.1 p10kw)
Number of 'GET' + noun hits (possession)	109 (6.1 p10kw)
Number of 'GET' + adjective hits (passive)	53 (2.9 p10kw)
Number of 'GET' + adjective hits (inchoative)	40 (2.2 p10kw)
Number of idiomatic 'GET' hits	63 (3.5 p10kw)

Table 5: figures for the American sub-corpus of the LOCNESS

A quick look at Table 5 shows how, even though the number of tokens and essays is quite similar to the British sub-corpus, the American one contains a much higher number of instances where *get* is used. It can express movement, especially through the phrases 'get home' (e.g. ICLE-US-SCU-0014.2: "*when my mom **gets home** she picks up my younger sisters from the baby-sitter's*") and 'get there' (e.g. ICLE-US-MICH-0037.1: "*I'm now here in Madrid. How did I **get there**?*") or in a metaphorical way (e.g. ICLE-US-MRQ-0003.1: "*it is simply impossible for the average person to **get inside the mind** of a criminal and see what he or she fears most*").

When followed by a noun phrase, it denotes a possession meaning, expressing either something received (e.g. ICLE-US-MRQ-0004.1: "*They said there was a need for greater cultural awareness and complained that administrators built these centers in places where they would not **get too much attention***"), something obtained (e.g. ICLE-US-MICH-0009.1: "*there neighborhoods don't **get financial support** and money is not spent in the stores near them*"), or also something caught, when referred to diseases (e.g. ICLE-US-PRB-0038.2: "*a case worker once reported a welfare mother who thought she had to*

quit work because her child ***got chicken pox***"; ICLE-US-MICH-0020.1: "*everyone was in a panic that their children would* ***get AIDS*** *simply from being around Ryan*").

When followed by an adjective, on the other hand, it forms either the passive (e.g. ICLE-US-SCU-0015.2: "*a young man did not* ***get treated*** *like that, he was basically able to do as he pleased*"), with the example 'GET caught' being particularly relevant, as it is present eight times (0.4 p10kw) (e.g. ICLE-US-IND-0014.1: "*besides the fact that you will eventually* ***get caught*** *doing a crime your punishment varies according to the type of crime or crimes that the individual has committed*") or it can take an inchoative meaning (e.g. ICLE-US-PRB-0026.1: "*they tend to* ***get frustrated*** *and then wish to relay this frustration through aggression*"), usually with the comparative form of the adjective here as well (e.g. ICLE-US-SCU-0007.2: "*there are a limited amount of teachers because the pay isn't good and rules are* ***getting stricter*** *concerning teacher's limitations*"; ICLE-US-PRB-0026.1: "*as he continues to talk about her, he should gradually* ***get more positive*** *in his characterization of this person*").

The meaning of causation can also be found in this sub-corpus, especially through the structure *get* + noun + to-infinitive (e.g. ICLE-US-SCU-0013.3: "*the justice system often resorts to desperate measures in order to* ***get the reporter to reveal*** *his/her sources*") or *get* + noun + adjective (e.g. ICLE-US-SCU-0012.3: "*the wild card* ***gets more cities involved*** *in the playoff race because it allows an additional team from any division in each league to participate in the playoffs*"), as well as the ingressive one (e.g. ICLE-US-SCU-0001.1: "*when fans* ***get to see*** *this every once in a while they will buy more tickets to the game and once they get into the stadium they will buy more merchandise to support their team*").

As regards the past form of the verb *got*, it has a higher number of hits in comparison with the British counterpart (42 hits, 2.3 p10kw), however American learners similarly do not produce often its use as a modal, since it is present only two times (0.1 p10kw) (ICLE-US-SCU-0002.4: "***you have got to remember*** *that we have some small children that ride with their parents on the track*", "*all amusement parks have to have rules,* ***you got to remember*** *that an amusement park is a place where families and friends can go and have fun*") as well as the form 'have got' to express possession, which is present only one time (0.05 p10kw) (ICLE-US-SCU-0002.3: "*but if both teams win,* ***you've got*** *a major dilemma*"). The American sub-corpus is the only one in this study which presents

instances of the past form *gotten* as well: however, it is not used more frequently than *got*, as it can be found only in 5 hits (0.2 p10kw) (e.g. ICLE-US-PRB-0030.1: "*I know that there has been a lot of progress in terms of segregation and discrimination, but it really couldn't have* ***gotten*** *much worse that it had been*").

The phrasal verbs found in this sub-corpus are similar in quantity and type to the British counterpart, and are precisely: 'get ahead', in its meaning of making progress (e.g. ICLE-US-SCU-0006.1: "*they spend their whole life trying to* ***get ahead*** *in a job and make a good living for the family*"); 'get at', in its sense of meaning (ICLE-US-IND-0021.1: "*so what am I* ***getting at****? Only that so-called feminists have not in and of themselves harmed the Women's movement*"); 'get away', in its sense of escaping (e.g. ICLE-US-PRB-0009.1: "*An-Mei lies about the nightmare she had in order to* ***get away*** *from a disasterous situation*"); 'get away with', in its sense of managing to do something that others would normally find unacceptable (e.g. ICLE-US-SCU-0010.4: "*the basic stereotypes are pink is for females, and blue is for males. Females can wear masculine colors and* ***get away with it****, but males can not wear pink, a feminine color, and* ***get away with it***"); 'get back', in its sense of returning (e.g. ICLE-US-MRQ-0030.1: "*the Government's supporting role will serve as sustenance and encouragement for it's people and allow them to* ***get back*** *into the tax base*"); 'get into' in its sense of being involved (ICLE-US-SCU-0007.2: "*teachers who really* ***get into*** *their work by going an extra mile to help students in need, should be rewarded*"); 'get off', in its sense of moving away (ICLE-US-SCU-0005.2: "*it would be a great a great opportunity for even more teenagers and adults to* ***get off*** *the streets*") or of getting out, however in a metaphorical way (ICLE-US-SCU-0012.4: "*I would like to say to the editors and publishers of today and to the newscasters, please* ***get off*** *this merry-go-round that you are on*"); 'get on', in its sense of boarding on a public transport, is also used metaphorically (ICLE-US-PRB-0004.1: "*Phillis Wheathey also tells African Americans to* ***get on*** *that "angelic train"*"); 'get out', in its sense of leaving (e.g. ICLE-US-SCU-0006.3: "*she is still receiving payments but is also receiving an education that may help her to someday* ***get out*** *of the welfare program*") and of escaping, referred to prisoners (ICLE-US-MRQ-0004.1: "*they demand blacks to* ***get out*** *of jail, stope using and selling drugs, and end the cycle of teenage pregnancies*"); 'get through', in its sense of finishing something (ICLE-US-SCU-0002.1: "*this unfortunately puts alot of pressure on men, as they aspire to* ***get through*** *college*") and of making someone

understand something (ICLE-US-MRQ-0001.1: "*because of this, they are much more effective in **getting through** to someone of the opposing argument"*); 'get together', in its sense of meeting (ICLE-US-SCU-0008.2: "*what the teams do is the managers **get together** and decide if they want you to play with them the next year*"), 'get to', in a metaphorical sense of reaching a place (ICLE-US-MICH-0012.1: "*positive aspects are those such as "Peace Talks" between nations; these would not take place as often as the do, if it took the President 1 week to **get to** where they were going*") and 'get up', in its sense of leaving the bed (e.g. ICLE-US-SCU-0012.4: "*why do they have to know and report when a person gets up in the morning, or when that person goes to bed at night?*"). As for idioms, there are six occurrences in this sub-corpus, thus twice as many as the British counterpart: 'get carried away' (ICLE-US-PRB-0029.1: "*it is much too easy to **get carried away** in aggresive feelings and end up killing someone*"), 'get in contact' (ICLE-US-IND-0018.1: "*there are key points to this in that a person has to **get in contact** with the right people, which takes time*"), 'get in the way' (ICLE-US-PRB-0031.1: "*even small things such as **getting in the way** of someone's goal can lead to the frustration which leads to aggressive act of murder*"), 'get on with life' (ICLE-US-MRQ-0046: "*here they intend to show the reader that the pain experienced by the family tends to become more of a psychological barrier to **getting on with their lives***"), 'get there' (e.g.ICLE-US-MRQ-0014.1: "*Ben Johnson, a world class sprinter, is an example of an athlete and role model striving to be the best and ended up using steroids to **get there***") and 'get the most out of something' (ICLE-US-SCU-0001.3: "*premarital sex is usually in a setting where both parties are trying to **get the most out of** the situation without giving of themselves totally*"). Additional[7] idioms found in the sub-corpus are 'get even', which means to take revenge (ICLE-US-MRQ-0003.1: "*it is simply based upon the human desire to **get even** and seek revenge*") and 'get it over with', which means to do an unpleasant task promptly (ICLE-US-MRQ-0023.1: "*but society would just as soon bury someone with AIDS alive, and **get it over with***").

[7] "Additional" referred to idioms means that they are not listed in the third chapter.

4.5.3 A comparison between British and American students

	British sub-corpus	American sub-corpus
Tokens	155,891	178,109
Number of 'get' hits	112 (7.1 p10kw)	305 (17.1 p10kw)

Table 6: figures for the two LOCNESS sub-corpora compared

As can be seen in Table 6, the American sub-corpus presents a higher frequency of use of the verb *get* and as for the specific meanings, all of them can be found in both corpora. Some observations that can be made are for example how in the American corpus the sense of possession is expressed more frequently through *got* and only once with *have got*, as Americans tend to omit the verb *have*, while for British learners is the opposite. Moreover, it can be seen how both learners produce a relevant quantity of phrasal verbs and idioms: in the British corpus their number is even higher than that of passives and inchoative structures (*get* + adjectives), however in the American corpus there are almost twice as many (63 hits, 3.5 p10kw versus 32 hits, 2.05 p10kw), with a peculiar high frequency of *get away with*, which in fact appears for a total of 10 hits (0.5 p10kw). In addition, the majority of phrasal verbs throughout the two sub-corpora is of the same typology (e.g. *get back*, *get on*, *get off*), while idioms, apart from *get on with one's life*, are all different from each other, which shows both a wider variety in native corpora and a difference in the most used expressions between the two variants of English.

4.6 A comparison between ICLE and LOCNESS and final observations

	ICLE	LOCNESS
Tokens	407,205	334,000
Number of 'GET' hits	795 (19.5 p10kw)	418 (12.5 p10kw)
Number of GET' + noun hits (possession)	275 (6.7 p10kw)	144 (4.3 p10kw)

Number of 'GET' + adjective hits (passive)	97 (2.3 p10kw)	69 (2.06 p10kw)
Number of 'GET' + adjective hits (inchoative)	124 (3.04 p10kw)	48 (1.4 p10kw)
Number of idiomatic 'GET' hits	194 (4.7 p10kw)	95 (2.8 p10kw)
Number of other uses of 'GET' hits	31 (0.7 p10kw)	36 (1.07 p10kw)

Table 7: figures of ICLE (Italian + German) and LOCNESS (British + American) corpora

The last comparison of this study examines learner versus native corpora: the normalized frequency scores show how Italian and German learners produced more examples of *get* than native British and American ones, a result which is congruent with the study of Ringbom (1998) mentioned in the previous chapter, as he came to the conclusion that some high-frequency verbs are found more frequently in non-native corpora. The author also argued that, however, non-native learners produced frequently the verb in question but mainly with its main meaning: this can be seen also in this study, as *get* is mostly used in its sense of obtaining or receiving something when followed by a noun phrase. However, Table 7 shows how the same pattern can be seen for L1 students too, which implies no relevant difference in use between the two categories of learners. Moreover, the fact that many synonyms of *get* can also be found in the learner corpus, and sometimes even more frequently (such as *become* or *obtain*), indicates that its use by non-native learners should not be explained in terms of 'safe choice', as for other high-frequency verbs, since they prove to know how to use alternatives as well.

A difference between the two categories of students which can be noted by taking into consideration the use of *get* plus adjectives only is that L2 students tend to produce it more with an inchoative meaning, while natives on the contrary through passives; furthermore, one last look at the other uses of *get*, which includes the ingressive and causation ones, shows that this is the only category where L1 students produced more instances, even though the difference between the two is minimal.

In relation to phrasal verbs, which are counted together with idioms under 'idiomatic GET hits', it can be seen how they are produced by both categories of learners in a considerable amount. According to the studies mentioned in the previous chapter, non-native learners

should present in their essays less examples, as it should be more difficult for them to learn how to use them and their various meanings; however it can be seen how they actually produce twice as many if compared to native speakers: this result also seems to contradict the fact that natives make extensive use of phrasal verbs.

However, since as seen above the single sub-corpora vary in many aspects, especially the Italian and German ones, a more precise confrontation can be made by looking at the individual figures for each category, provided in Table 8:

	Italian sub-corpus	German sub-corpus	British sub-corpus	American sub-corpus
Tokens	209, 524	197,681	155,891	178,109
Number of 'GET' hits	198 (9.4 p10kw)	597 (30.2 p10kw)	112 (7.1 p10kw)	305 (17.1 p10kw)
Number of 'GET' + noun hits (possession)	87 (4.1 p10kw)	188 (9.5 p10kw)	35 (2.2 p10kw)	109 (6.1 p10kw)
Number of 'GET' + adjective hits (passive)	48 (2.2 p10kw)	49 (2.4 p10kw)	16 (1.02 p10kw)	53 (2.9 p10kw)
Number of 'GET' + adjective hits (inchoative)	26 (1.2 p10kw)	98 (4.9 p10kw)	8 (0.5 p10kw)	40 (2.2 p10kw)
Number of idiomatic 'GET' hits	26 (1.2 p10kw)	168 (8.5 p10kw)	32 (2.05 p10kw)	63 (3.5 p10kw)
Number of other uses of 'GET' hits	2 (0.09 p10kw)	29 (1.4 p10kw)	12 (0.7 p10kw)	24 (1.3 p10kw)

Table 8: figures for the individual sub-corpora of ICLE and LOCNESS

One final look at all the sub-corpora investigated in this study in Table 8 confirms how the possession meaning is the most frequent in each single one; moreover, it is also possible to note how only in the Italian sub-corpus the number of passive structures is higher than inchoative ones but also than idiomatic forms, while for all the others idioms and phrasal verbs represent the second most frequent category.

With reference to Coto-Villalibre's study about *get* (2016) mentioned in 3.4, it is possible to detect some similarities, as Italian and German learners too use the verb more often in adjectival constructions than in idiomatic ones. However, the author does not take into consideration the constructions of the verb with noun phrases, which are the most found in this study: further research could investigate this aspect too, in order to see if Hong Kong learners present the same pattern of use of the verb.

On a concluding note, this study has found that generally Italian and German learners of English know how to use the verb *get*, through its main meanings when followed by nouns and adjectives, but also through some idioms and phrasal verbs. A comparison between the two sub-corpora shows how German learners have a wider knowledge of its uses, since they also produce some examples of *get* with an ingressive meaning, which is missing in Italian essays; moreover, the fact that many more examples could be found for each sense, and produced by different students, indicates that many of them know how to use the various structures, meaning that generally this category of learners have a good knowledge of the verb. The L1 corpus LOCNESS, which should present many more uses and instances as based on previous studies analysed, shows instead a lower frequency of *get*, with even half examples of each sense as compared to non-native learners ones.

As for the errors found in the ICLE essays, German and Italian learners do not show particular differences in terms of quantity and variety of them, as they both produce some due to interference with their L1 and some other syntactical ones.

Therefore it can be suggested that both learners could take advantage from working with corpora, in order to see the contexts in which high-frequency verbs are found and to acquire more familiarity with their many uses.

CONCLUSIONS

The aim of this thesis was to investigate how Italian and German learners of English use the verb *get*, by looking at two components of a learner corpus and in order to being able to suggest aimed tasks for the two categories.

I provided thus a general description of corpora at the beginning, focusing especially on the ways in which students and teachers can benefit from them in their work in the context of languages, and then I provided information about learner corpora only in the second chapter, namely their pedagogical uses through dictionaries, coursebooks and data-driven learning and previous research which analysed determinate characteristics of L2 learners, often comparing two or more categories. The studies reported regarded especially Italian and German learners, in order to have an idea of any difficulties they experience while producing in the English language and to already detect any similarities or differences between them.

Since the main topic of my research has been the verb *get*, before describing my study I provided a description of it in the third chapter: I focused first on the many uses which it can have and the dominant patterns which it can form, especially through phrasal verbs, and then I provided studies about these in order to analyse the factors which influence learners' production and above all the difficulties they face. I also retrieved one study about the verb itself based on a learner corpus, namely the Hong Kong component of the ICLE, which made possible to have a first look about how one category of L2 learners of English use the verb, especially any form of underuse, overuse and misuse.

In the last chapter I carried out my own study about Italian and German learners and their use of the verb: for the data I used the two sub-components of the ICLE learner corpus, I-ICLE and G-ICLE, and the concordance software *AntConc* to retrieve all the examples of *get* produced by the learners. After analysing both categories and going through all the meanings and mistakes found, I made a comparison between the two and made some suggestions about tasks that could be developed in the classroom; in the last part I then used the same procedure with the L1 corpus LOCNESS, made up of essays written by British and American students, and made a final comparison between all four categories of students, in order to see in which aspects L2 students differ more from L1 ones when producing the verb.

The main research questions which I was interested to investigate were if the two categories of learners tend to make the same mistakes when using *get*, and if with the same frequency, if they both use the verb with more than just its main meaning, namely also through phrasal verbs and idioms and through more complicated structures such as the causation and ingressive one, and finally if there are major differences when comparing their data with those of a L1 corpus; based on the results found, the final aim of this thesis was then to being able to suggest tasks aimed at learners and teachers of these two languages.

I first came to the conclusion that German learners use the verb relevantly more frequently than Italian ones: in their production all the meanings analysed can be found, with many examples produced by more than just a few students, while Italian learners do not use at all the ingressive meaning and for other uses, namely the causation and the movement ones, only a couple of examples could be found: these results showed therefore how they both use more than just the main meaning of *get*, however Italian learners show more difficulties, or probably a lack of knowledge, with the more complicated ones.

On the other hand, some similarities can also be seen: in fact, for both categories *get* is used more frequently in its possession meaning when followed by a noun phrase; moreover, as regards the errors, both learners present difficulties as to when the verb is followed by an infinite or an -ing form and others due to wrong literal translations from their L1, and for this reason it can be stated that mistakes are only partially the same between the two.

Regarding the comparison between L2 and L1 students, an interesting result was that the former produced more instances than the latter: this contradicted the idea that natives make a much wider use of words than learners, but also confirmed the results of the study of Ringbom (1998) that some high-frequency verbs are overused by L2 learners. Moreover, another striking result was that both categories of students followed the same pattern of use, which means that the most frequent uses for both were the possession and the idiomatic ones: a much wider use was in fact expected from L1 students, especially of phrasal verbs, since for L2 ones it is often difficult to remember and produce them.

Only a partial comparison could then be made with the Hong Kong learners studied by Coto-Villalibre (2016), as the author does not consider noun constructions of *get*: however, it was possible to see how adjectival constructions, meaning passive and inchoative ones,

were more used than idiomatic ones for all the three categories of learners.

From these results, I suggest that corpora could be a useful aid in the English studies of Italian and German learners, as they had a rather high average age and still presented some syntactical errors while producing the verb. Concordance lines taken from L1 corpora can be compared to sentences produced in their own essays, in order for them to see the correct versions and learn from their mistakes. In particular, the category of phrasal verbs and idioms showed the most notable differences, as Italian learners produced a small quantity of them while German ones on the contrary a very broad one, however with a high number of mistakes. Different exercises could therefore be suitable for them: Italian learners could make use of concordances of phrasal verbs and idioms which contain many examples where they have to indicate the meaning, starting from the main ones formed by the most common high-frequency verbs, in order to enlarge their knowledge; German learners could instead be given exercises where parts of phrasal verbs and idioms are taken out, in order to make practice on how they are structured, since they do not show problems in terms of quantity but rather on the correct way to produce them.

In my research, however, I considered only one high-frequency verb and only two categories of L2 learners: for this reason, it would be interesting both to carry out studies with the same categories of students about other frequent verbs and to carry out studies about *get* considering other L2 learners. Other interesting common high-frequency verbs to study could be *make* or *take*, which as well as *get* have many meanings and form many phrasal verbs and idioms: it would be thus interesting to see if the same peculiarities can be found and if Italian and German learners need more practice and exercise with *get* only or with high-frequency verbs in general. Other L2 learners to consider could be the others included in the ICLE, in order to have a general idea of the use of the verb in the same period of time.

It would also be interesting if further research could investigate the same verb and the same categories of learners through other corpora, in order to see if these general results can be confirmed: for example, in a spoken corpus it could be possible to better analyse the informal register where *get* is mostly found, or in addition more recent corpora could show any new meaning of the verb, since after 2009 (when ICLE was released) there has been quite a linguistic change due to the popularity of social media.

BIBLIOGRAPHY

Aijmer, Karin (2002). Modality in advanced Swedish learners' written interlanguage, *in* S. Granger, J. Hung & S. Petch-Tyson (eds.) *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins, pp. 55-76.

Altenberg, Bengt & Tapper, Marie (1998). The use of adverbial connectors in advanced Swedish learners' written English, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 80-93.

Aston, Guy (1995). Corpora in language pedagogy: Matching theory and practice, *in* G. Cook and B. Seidlhofer (eds) *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson*. Oxford: Oxford University Press, pp. 257-270.

Aston, Guy (1997). Enriching the Learning Environment: Corpora in ELT, *in* A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.) *Teaching and Language Corpora.* London: Longman, pp. 51-64.

Becker, Anthony (2014). Avoidance of English Phrasal Verbs: Investigating the Effect of Proficiency, Learning Context, Task Type, and Verb Type. *Asian Journal of English Language Teaching, 24*, pp. 1-33.

Bernardini, Silvia (2004). Corpora in the Classroom: an Overview and some Reflections on Future Developments, *in* Sinclair, John (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: Benjamins, pp. 15-36.

Biber, Douglas & Johansson, Stig & Leech, Geoffrey & Conrad, Susan and Finegan, Edward (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Bolinger, Dwight (1971). *The Phrasal Verb in English*. Cambridge, MA: Harvard University Press.

Boulton, Alex (2010). Data-Driven Learning: On Paper, in Practice, *in* T. Harris & M. Moreno Jaén (eds.) *Corpus Linguistics in Language Teaching*. Bern: Peter Lang, pp. 17-52.

Boulton, Alex (2012). Hands-on / hands-off: Alternative approaches to data-driven learning, *in* J. Thomas & A. Boulton (eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press, pp. 153-168.

Boulton, Alex (2019a). First contact with language corpora: perspectives from students, *in* F. Meunier, J. Van de Vyver, L. Bradley & S. Thouësny (eds.) *CALL and complexity – short papers from EUROCALL 2019.* pp. 51-56.

Boulton, Alex (2019b). Data-driven learning for younger learners: Obstacles and Optimism, *in* P. Crosthwaite (ed.) *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge, pp. 14-20.

Boulton, Alex & Cobb, Tom (2017). Corpus use in language learning: A meta-analysis. *Language learning, 67 (2)*, pp. 348-393.

Boulton, Alex & De Cock, Sylvie (2017). Dictionaries as aids for language learning, *in* P. Hanks & G.M. de Schryver (eds.) *International Handbook of Lexis and Lexicography*. New York: Springer.

Braun, Sabine (2010). Getting Past 'Groundhog Day': Spoken Multimedia Corpora for Student-centred Corpus Exploration, *in* T. Harris & M. Moreno Jaén (eds.) *Corpus Linguistics in Language Teaching*. Bern: Peter Lang, pp. 75- 97.

Breckle, Margit & Zinsmeister, Heike (2013). L1 transfer versus fixed chunks: A learner corpus-based study of L2 German, *in* S. Granger, G. Gilquin & F. Meunier (eds.) *Twenty years of Learner Corpus Research: Looking back, Moving ahead*. Corpora and Language in Use – Proceedings 1, Louvain-la-Neuve: Presses universitaires de Louvain, pp. 25-35.

Breyer, Yvonne Alexandra (2011). *Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges*. Frankfurt am Main: Peter Lang.

Callies, Marcus (2009). "What's even more alarming is…" - a contrastive learner-corpus study of *what*-clefts in advanced German and Polish L2 writing, *in* M. Wysocka (ed.) *On Language Structure, Acquisition and Teaching*. *Studies in honour of Janusz Arabski on the occasion of his 70th birthday*. Katowice: Wydawnictwo Uniwersytetu Slaskiego, pp. 283-292.

Callies, Marcus (2015). Using Learner Corpora in Language Testing and Assessment: Current Practice and Future Challenges, *in* E. Castello, K. Ackerley & F. Coccetta (eds.) *Studies in Learner Corpus Linguistics*. Bern: Peter Lang, pp. 21-35.

Callies, Marcus & Zaytseva, Ekaterina (2013). The *Corpus of Academic Learner English* (CALE) – A new resource for the study and assessment of advanced language proficiency, *in* S. Granger, G. Gilquin & F. Meunier (eds.) *Twenty Years of Learner Corpus Research: Looking back, Moving ahead*. Corpora and Language in Use – Proceedings 1, Louvain-la-Neuve: Presses universitaires de Louvain, pp. 49-59.

Castello, Erik (2013). Integrating learner corpus data into the assessment of spoken interaction in English in an Italian university context, *in* S. Granger, G. Gilquin & F. Meunier (eds.) *Twenty years of Learner Corpus Research: Looking back, Moving ahead*. Corpora and Language in Use – Proceedings 1, Louvain-la-Neuve: Presses universitaires de Louvain, pp. 61-74.

Castello, Erik (2017). *It*-extraposition constructions in Italian EFL academic writing: A longitudinal study based on the Italian component of the LONGDALE project, *in* P. de Haan, S. van Vuuren & R. de Vries (eds.) *Language, Learners and Levels: Progression and Variation*. Corpora and Language in Use – Proceedings 3, Louvain-la-Neuve: Presses universitaires de Louvain, pp. 153-180.

Chambers, Angela (2010). Contrastive Language Data: From Translation Studies to Language Learning and Teaching, *in* T. Harris & M. Moreno Jaén (eds.) *Corpus Linguistics in Language Teaching*. Bern: Peter Lang, pp. 99-118.

Chen, Meilin (2013). Phrasal verbs in a longitudinal learner corpus: Quantitative Findings, *in* S. Granger, G. Gilquin & F. Meunier (eds.) *Twenty years of Learner Corpus Research: Looking back, Moving ahead*. Corpora and Language in Use – Proceedings 1, Louvain-la-Neuve: Presses universitaires de Louvain, pp. 89-101.

Chujo, Kiyomi & Utiyama, Masao & Miura, Shinji (2006). Using a Japanese-English parallel corpus for teaching English vocabulary to beginning-level students. *English Corpus Studies 13*: pp. 153-172.

Chujo, Kiyomi & Oghigian, Kathryn (2012). DDL for EFL beginners: A report on students gains and views on paper-based concordancing and the role of L1, *in* J. Thomas & A. Boulton (eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press, pp. 169-182.

Cobb, Tom (1997). Is there any measurable learning from hands-on concordancing? *System, Vol. 25, No. 3*, pp. 301-315.

Coto-Villalibre, Eduardo (2016). A Look at Participial Constructions with Get in Hong Kong English, *in* M. J. López-Couso, B. Mendez-Naya, P. Núñez-Pertejo, & I. M. Palacios-Martínez (eds.), *Corpus linguistics on the move: Exploring and understanding English through corpora*, vol 79, Brill/Rodopi, pp. 204-226.

Cresswell, Andy (2015). Representing Learner English in a Specialized Corpus: Genre and Proficiency Level in the Advanced Learner Corpus of Argumentative Student Essays (ALCASE), *in* E. Castello, K. Ackerley & F. Coccetta (eds.) *Studies in Learner Corpus Linguistics*. Bern: Peter Lang, pp. 297-315.

Cullen, Kay & Hands, Penny & McGovern, Una & Wright, John (2003). *Phrasal Verbs – a Practical Guide to Common Phrasal Verbs from British and American English*. Milano: Ulrico Hoepli Editore.

Dagut, Menachem & Laufer, Batia (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition 7*, pp. 73-79.

De Cock, Sylvie & Granger, Sylviane & Leech, Geoffrey & McEnery, Tony (1998). An automated approach to the phrasicon of EFL learners, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 67-79.

De Cock, Sylvie & Granger, Sylviane (2004). Computer Learner Corpora and Monolingual Learners' Dictionaries: the Perfect Match, *in* W. Teubert and M. Mahlberg (eds.) *The Corpus Approach to Lexicography*. Special issue of Lexicographica 20: pp. 72-86.

De Cock, Sylvie & Paquot, Magali (2010). The monolingual learners' dictionary as a productive tool: the contribution of learner corpora, *in*: M.C. Campoy & B. Belles-Fortuno & M.-L. Gea-Valor (eds.) *Corpus-Based Approaches to English Language Teaching*, Continuum. London: New York, pp. 195-204.

Díaz-Negrillo, Ana & Thompson, Paul (2013). Learner corpora: Looking towards the future, *in* A. Díaz-Negrillo, N. Ballier & P. Thompson (eds.) *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins, pp. 9-29.

Donesch-Jezo, Ewa (2013). Using Language Corpus in Teaching Foreign Language Vocabulary. *International Multidisciplinary e – Journal*, pp. 11-25.

Frankenberg-Garcia, Ana (2012). Integrating corpora with everyday language teaching, *in* J. Thomas & A. Boulton (eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press, pp. 37-53.

Gairns, Ruth & Redman, Stuart (2011a). *Idioms and Phrasal Verbs – Intermediate*. Oxford: Oxford University Press.

Gairns, Ruth & Redman, Stuart (2011b). *Idioms and Phrasal Verbs – Advanced*. Oxford: Oxford University Press.

Gallina, Francesca (2013). The Lexicon of Spoken Italian by Foreigners: A study on the acquisition of vocabulary by L2 Italian learners between measures of lexical richness and lexical fields, *in* S. Granger, G. Gilquin & F. Meunier (eds.) *Twenty years of Learner Corpus Research: Looking back, Moving ahead*. Corpora and Language in Use – Proceedings 1, Louvain-la-Neuve: Presses universitaires de Louvain, pp. 179-195.

Gardner, Dee & Davies, Mark (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. TESOL Quarterly 41 (2), pp. 339-360.

Gavioli, Laura (1997). Exploring Texts through the Concordancer: Guiding the Learner, *in* A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.) *Teaching and Language Corpora*. London: Longman, pp. 83-99.

Gavioli, Laura (2009). Corpus Analysis and the Achievement of Learner Autonomy in Interaction, *in* L. Lombardo (ed.) *Using Corpora to Learn about Language and Discourse*. Bern: Peter Lang, pp. 39-71.

Geist, Monika & Hahn, Angela (2012). Using a corpus for written production: A classroom study, *in* J. Thomas & A. Boulton (eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press, pp. 123-133.

Gerckens, Caroline & Gans, Anne (2015). I didn't really *understood what it was about, but it really *made fun: A longitudinal Corpus-based Study of Tense/Aspect and High-frequency Verbs in Learner English, *in* E. Castello, K. Ackerley & F. Coccetta (eds.) *Studies in Learner Corpus Linguistics*. Bern: Peter Lang, pp. 143-157.

Gillard, Patrick & Gadsby, Adam (1998). Using a learners' corpus in compiling ELT dictionaries, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 159-171.

Gilmore, Alex (2009). Using online corpora to develop students' writing skills. *ELT Journal 63(4)*: pp. 363-372.

Götz, Sandra (2012). Temporal fluency variables in native and non-native English speech: Corpus findings and language-pedagogical implications, *in* J. Thomas & A. Boulton (eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press, pp. 229-243.

Gouverneur, Céline (2008). The phraseological patterns of high-frequency verbs in advanced English for general purposes: A corpus-driven approach to EFL textbook analysis, *in* F. Meunier and S. Granger (eds.) *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, pp. 223-246.

Granath, Solveig (2009). Who benefits from learning how to use corpora? *in* K. Aijmer (ed.) *Corpora and Language Teaching*. Amsterdam: John Benjamins, pp. 47-65.

Granger, Sylviane (1997). On identifying the syntactic and discourse features of participle clauses in academic English: native and nonnative writers compared, *in*: J. Aarts, I. de Mönnink, and H. Wekker (eds.) *Studies in English Language and Teaching*. Rodopi: Amsterdam & Atlanta, pp. 185-198.

Granger, Sylviane (1998). The computer learner corpus: a versatile new source of data for SLA research, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 3-18.

Granger, Sylviane (1999). Use of tenses by advanced EFL learners: Evidence from an error-tagged computer corpus, *in* H. Hasselgard & S. Oksefjell (eds.) *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi, pp. 191-202.

Granger, Sylviane (2002). A bird's eye view of learner corpus research, *in* S. Granger, J. Hung & S. Petch-Tyson (eds.) *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins, pp. 3-36.

Granger, Sylviane (2004). Computer learner corpus research: current status and future prospects', *in* U. Connor and Th. A. Upton (eds.) *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi, pp. 123–145.

Granger, Sylviane (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation, *in* K. Aijmer (ed.) *Corpora and Language Teaching*. Amsterdam: John Benjamins, pp. 13-32.

Granger, Sylviane & Rayson, Paul (1998). Automatic profiling of learner texts, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 119-131.

Granger, Sylviane & Tribble, Chris (1998). Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 199-209.

Granger, Sylviane & Paquot, Magali (2009). Lexical verbs in academic discourse: a corpus-driven study of learner use, *in* M. Charles, D. Pecorari & S. Hunston (eds.) *Academic Writing. At the Interface of Corpus and Discourse*. London & New York: Continuum, pp. 193-214.

Granger, Sylviane & Dagneaux, Estelle & Meunier, Fanny & Paquot, Magali (2009). *International Corpus of Learner English v2*. Louvain-la-Neuve: Presses universitaires de Louvain.

Gronemeyer, Claire (1999). On deriving complex polysemy: the grammaticalization of *get*. *English Language and Linguistics 3 (1)*. Cambridge University Press: pp. 1-39.

Hulstijn, Jan & Marchena, Elaine (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition 11*, pp. 241-255.

Hunston, Susan (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, Susan (2009). The usefulness of corpus-based descriptions of English for learners: The case of relative frequency, *in* K. Aijmer (ed.) *Corpora and Language Teaching*. Amsterdam: John Benjamins, pp. 141-154.

Jones, Christian & Byrne, Shelley & Halenko, Nicola (2018). *Successful Spoken English: Findings from Learner Corpora*. Routledge.

Johns, Tim (1991). Should you be persuaded – Two samples of data-driven learning materials, *in* T. Johns & P. King (eds), *Classroom Concordancing. English Language Research Journal. 4*, pp. 1-16.

Johns, Tim (2002). Data-Driven Learning: the Perpetual Challenge, *in* B. Ketteman & G. Marko (eds.) *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, pp. 107-118.

Knowles, Gerry (1997). Using Corpora for the Diachronic Study of English, *in* A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.) *Teaching and Language Corpora.* London: Longman, pp. 195-210.

Leech, Geoffrey (2011). Frequency, corpora and language learning, *in* F. Meunier, S. De Cock, G. Gilquin and M. Paquot (eds) *A Taste for Corpora: In honour of Sylviane Granger*, John Benjamins Publishing Company, pp. 7-27.

Leech, Geoffrey, Rayson, Paul & Wilson, Andrew (2001). *Word frequencies in written and spoken English*. Harlow: Longman.

Lennon, Paul (1996). Getting 'easy' verbs wrong at the advanced level. *International Review of Applied Linguistics 34* (1), pp. 23-36.

Liao, Yan & Fukuya, Yoshinori (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning, 54*, pp. 193–226.

Liu, Dilin (2011). The most-frequently used English phrasal verbs in American and British English: A multi-corpus examination. *TESOL Quarterly 45 (4)*, pp. 661-688.

Lorenz, Gunter (1998). Overstatement in advanced learners'writing: stylistic aspects of adjective intensification, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 53-66.

Lorenz, Gunter (2009). The Status of English in Germany, in S. Granger, E. Dagneaux, F. Meunier and M. Paquot (eds.) *International Corpus of Learner English v2*. Louvain-la-Neuve: Presses universitaires de Louvain. pp. 130-138.

López Pérez, Sidoni & Benali Taouis, Hanane (2019). Analysis of noun (direct object) collocations with the high-frequency verb DO by Spanish students in an online learner corpus, *in Complutense Journal of English Studies 27*, pp. 99-120.

McArthur, Tom & Atkins, Beryl (1974). *Dictionary of English Phrasal Verbs and their Idioms*. London: Collins.

McCarthy, Michael & O'Dell, Felicity (2004). *English Phrasal Verbs in Use*. Cambridge, England: Cambridge University Press.

Meunier, Fanny & Gouverneur, Céline (2007). The treatment of phraseology in ELT textbooks, *in* E. Hidalgo, L. Quereda & J. Santana (eds.) *Corpora in the Foreign Language Classroom. Selected papers from the Sixth International Conference on Teaching and Language Corpora* (TaLC 6). University of Granada, Spain, 4-7 July, 2004. Amsterdam: Rodopi, pp. 119-139.

Meunier, Fanny (2016). Learner Corpora and pedagogic applications, *in* L. Murray & F. Farr (eds.) *The Routledge Handbook of Language Learning and Technology*. Routledge, pp. 376-387.

Milton, John (1998). Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 186-198.

Mindt, Dieter (1997). Corpora and the Teaching of English in Germany, *in* A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.) *Teaching and Language Corpora.* London: Longman, pp. 40-50.

Mukherjee, Joybrato (2009). The grammar of conversation in advanced spoken learner English, Learner corpus data and language-pedagogical implications, *in* K. Aijmer (ed.) *Corpora and Language Teaching*. Amsterdam: John Benjamins, pp. 203-230.

Nesselhauf, Nadja (2004). Learner corpora and their potential for language teaching, *in* Sinclair, John (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: Benjamins, pp. 125-152.

Nesselhauf, Nadja (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.

O'Dell, Felicity (1997). Incorporating vocabulary into the syllabus, *in* N. Schmitt & M. McCarthy (eds.) *Vocabulary: Description, acquisition and pedagogy*. Oxford: Oxford University Press, pp. 258-278.

O'Keeffe, Anne & McCarthy, Michael & Carter, Ronald (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.

Pérez-Paredes, Pascual (2010). Corpus Linguistics and Language Education in Perspective: Appropriation and the Possibilities Scenario, *in* T. Harris & M. Moreno Jaén (eds.) *Corpus Linguistics in Language Teaching*. Bern: Peter Lang, pp. 53-73.

Prat Zagrebelsky, Maria Teresa (2004). *Computer Learner Corpora: Theoretical Issues and Empirical Case Studies of Italian Advanced EFL Learners' Interlanguage*. Alessandria: Edizioni dall'Orso.

Prat Zagrebelsky, Maria Teresa (2009). The Status of English in Italy, in S. Granger, E. Dagneaux, F. Meunier and M. Paquot (eds.) *International Corpus of Learner English v2*. Louvain-la-Neuve: Presses universitaires de Louvain. pp. 139-143.

Quirk, Randolph & Greenbaum, Sidney & Leech, Geoffrey and Svartvik, Jan (1985). A comprehensive grammar of the English language. London, Longman.

Reznicek, Marc & Lüdeling, Anke & Hirschmann, Hagen (2013). Competing target hypotheses in the Falko corpus: A flexible multi-layer corpus architecture, *in* A. Díaz-Negrillo, N. Ballier & P. Thompson (eds.) *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins, pp. 101-123.

Ringbom, Håkan (1998). Vocabulary frequencies in advanced learner English: a cross-linguistic approach, *in* S. Granger (ed.) *Learner English on Computer*. London: Longman. pp. 41-52.

Rogatcheva, Svetla (2012). Measuring learner (mis)use: Tense and aspect errors in the Bulgarian and German components of ICLE, *in* J. Thomas & A. Boulton (eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press, pp. 258-272.

Römer, Ute (2004). Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching, *in* G. Aston, S. Bernardini & D. Stewart (eds.) *Corpora and Language learners*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 151-168.

Römer, Ute (2009). Corpus research and practice: What help do teachers need and what can we offer? *in* K. Aijmer (ed.) *Corpora and Language Teaching*. Amsterdam: John Benjamins, pp. 83-98.

Schmitt, Norbert & Redwood, Stephen (2011). Learner knowledge of phrasal verbs: A corpus-informed study, *in* F. Meunier, S. De Cock, G. Gilquin and M. Paquot (eds.) *A Taste for Corpora: In honour of Sylviane Granger*, Amsterdam: John Benjamins, pp. 173-207.

Sealey, Alison & Thompson, Paul (2004). What do you call the dull words? Primary school children using corpus-based approaches to learn about language. *English in Education, 38*(1): pp. 80-91.

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, John (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Sinclair, John & Renouf, Antoinette (1988). A lexical syllabus for language learning, *in* R. Carter & M. McCarthy (eds.) *Vocabulary and language teaching*. Harlow: Longman, pp. 140-158.

Stubbs, Michael (1995). Corpus evidence for norms of lexical collocation, *in* G. Cook and B. Seidlhofer (eds) *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson*, Oxford: Oxford University Press, pp. 245-256.

Teubert, Wolfgang (2004). Language and corpus linguistics, *in* M.A.K. Halliday & W. Teubert & C. Yallop and A. Čermáková, *Lexicology and Corpus Linguistics. An Introduction*. London: Continuum, pp. 73-112.

Teubert, Wolfgang & Čermáková, Anna (2004). Directions in corpus linguistics, *in* M.A.K. Halliday & W. Teubert & C. Yallop and A. Čermáková, *Lexicology and Corpus Linguistics. An Introduction*. London: Continuum, pp. 113-165.

Tono, Yukio (2003). Learner corpora: Design, development and applications, *in* D. Archer, P. Rayson, A. Wilson & T. McEnery (eds.) *Proceedings of the Corpus Linguistic 2003 Conference*, pp. 800-809.

Tribble, Chris & Jones, Glyn (1997). *Concordances in the classroom: A Resource Book for Teachers.* Houston: Athelstan.

Viberg, Åke (2002). Polysemy and disambiguation cues across languages: The case of Swedish *få* and English *get*, i*n* B. Altenberg and S. Granger (eds.) *Lexis in Contrast. Corpus-based approaches*. Amsterdam: Benjamins, pp. 119-150.

Wiemeyer, Leonie (2017). Intertextuality in L2 academic writing: The use of paraphrases by German learners of English, *in* P. De Haan, S. Van Vuuren & R. De Vries (eds.) *Language, Learners and Levels: Progression and Variation*. Corpora and Language in Use – Proceedings 3, Louvain-la-Neuve: Presses universitaires de Louvain, pp. 45-68.

Zhou, Xiujuan (2016). A corpus-based study on high frequency verb collocations in the case of “Have”. International Forum of Teaching and Studies 12(1): pp. 42-50.

Riassunto in italiano

Il presente elaborato è uno studio sull'uso del verbo inglese *get* da parte di studenti della lingua inglese italiani e tedeschi basato su due corpora delle rispettive lingue: essi consistono in collezioni di testi, principalmente di carattere argomentativo, prodotti da determinate categorie di individui e creati con il fine di studiare le particolarità e regolarità della lingua in questione.

Nello specifico, la tesi è suddivisa in quattro capitoli: il primo tratta dei corpora in generale e, oltre a cosa sono, viene descritto anche come vengono classificati e come l'avvento della tecnologia negli ultimi decenni ha cambiato radicalmente il modo in cui possono essere utilizzati. Essi infatti possono essere diversificati in base a numerosi criteri: possono essere formati da testi scritti oppure da trascrizioni di produzioni orali (*spoken corpora*), che a loro volta possono essere di un registro formale o informale; possono essere formati da un numero ridotto o al contrario elevato di testi, oppure anche solamente da parti di essi, in base alla loro tipologia; possono riguardare un ristretto numero di studenti in un lungo arco di tempo, con il fine di studiare gli eventuali cambiamenti con l'avanzare degli studi, oppure possono riguardare più categorie di studenti in un periodo specifico, con il fine di studiare principalmente le differenze e similitudini tra di loro; inoltre possono riguardare studenti madrelingua della lingua in questione (*native corpora*), di madrelingua diversa (*learner corpora*), oppure di entrambe le categorie, con gli stessi testi nelle due lingue disposti in modo parallelo per facilitare un immediato confronto visivo (*parallel corpora*).

L'apporto principale dato dall'avvento della tecnologia riguarda la metodologia in cui i corpora possono venire letti e analizzati: infatti, prima dell'invenzione dei computer, compilare un corpus significava scrivere manualmente ogni singolo testo o parte di esso ed effettuare delle ricerche comportava dover leggere intere collezioni di dati; al giorno d'oggi, invece, oltre a poter collezionare molti più testi di maggiore grandezza, si può anche usufruire di software che permettono di ricercare determinate parole o strutture e di ottenere i risultati in un modo veloce e che facilita notevolmente lo studio e il confronto dei dati. Questi possono essere riprodotti attraverso *frequency lists*, tramite le quali è possibile visualizzare la frequenza di una o più parole nel corpus, e stabilire perciò quanto

esse siano importanti nella lingua o nel tema specifico considerati, in base al tipo di ricerca e al corpus che si utilizza; un altro procedimento di ricerca è attraverso le *concordance lines*, che tramite un software permettono di visualizzare tutti gli esempi di una parola o gruppo di parole all'interno del corpus: tramite *KWIC* (*key word in context*), che è il più utilizzato, è possibile visualizzare anche le intere frasi da cui gli esempi sono stati tratti, permettendo perciò di vedere i vari usi in cui quel termine viene espresso in base al contesto in cui esso si trova; spesso questo tipo di ricerca permette di individuare le cosiddette *collocations*, che sono combinazioni di una o più parole che però formano un'unica unità lessicale, il cui significato non è derivabile dalle singole parti che le formano.

Nella descrizione dei corpora, mi soffermo in particolare su come queste loro caratteristiche possono essere di beneficio per studenti e professori nel contesto dell'insegnamento delle lingue: quest'ultimo anno infatti ho potuto sperimentare entrambi i due punti di vista, poiché ho frequentato il mio ultimo anno di università e ho iniziato ad insegnare inglese ad alcuni studenti di scuole medie; è stato perciò stimolante e utile investigare come i corpora possono essere usati nell'ambiente educativo da entrambe le prospettive. Per quanto riguarda le *frequency lists*, disporre di liste con le parole più frequenti in un corpus può essere un punto di partenza per gli insegnanti per capire che cosa è più rilevante in una lingua e quindi cosa è fondamentale che gli studenti imparino in un primo momento; le *concordance lines* inoltre possono essere di aiuto per la compilazione di dizionari e libri di testo, in particolare nella decisione degli esempi da riprodurre a seguito del termine ricercato nei primi o nei dialoghi ed esercizi nei secondi. Uno studio condotto da Römer (2008) mostra infatti come la maggior parte dei libri di testo in Germania presenti dialoghi inventati, il che non porta alcun beneficio agli studenti, mentre se si considerassero esempi tratti da *spoken corpora* essi potrebbero già prendere dimestichezza con il registro informale della lingua e vedere come spesso singole parole vengono usate diversamente in diversi contesti.

Dati tratti dai corpora, adattati allo specifico livello di educazione degli studenti, possono inoltre anche essere usati direttamente in classe e questa procedura è denominata *data-driven learning*, termine coniato da Tim Johns: se presi singolarmente, possono infatti essere utilizzati per spiegare un determinato argomento o struttura, come mostra lo studio di Johns (1991) dove l'autore ha utilizzato *concordance lines* dei verbi *convince* e

persuade per analizzarne le differenze; se messi a confronto con esempi prodotti dagli studenti di un'altra madrelingua, gli stessi dati possono essere utili anche per sottolineare eventuali errori o similitudini tra le due lingue.

Oltre a mostrare esempi ricavati dai corpora, gli insegnanti possono anche modificarli creando esercizi di vario tipo: ad esempio, per insegnare le preposizioni corrette che seguono alcuni verbi o aggettivi (es. *think about*, *interested in*) essi possono mostrare più *concondarce lines* del verbo o aggettivo in questione omettendo la preposizione, che sarà compito degli studenti poi aggiungere. Questi ultimi possono inoltre anche essere guidati nell'interpretare da soli dei dati particolari, come nel caso descritto da Gavioli (2009), dove hanno analizzato un corpus formato da recensioni cinematografiche per individuare le espressioni più comunemente utilizzate: essi sono arrivati alla conclusione che i critici utilizzano valutazioni differenti per gli attori in base al loro genere e hanno perciò agito anche da ricercatori, avendo osservato e studiato la lingua in questione. Questa procedura rappresenta un vantaggio per gli studenti poiché rinforza la loro autonomia, permette loro di collaborare non solo con altri studenti ma anche con gli insegnanti stessi, che vengono perciò visti in un'ottica meno autoritaria, e conferisce loro più motivazione: feedback di numerosi studi riguardanti esercizi di questo tipo mostrano infatti non solo come siano effettivamente utili per gli studenti, ma anche come essi si sentano più soddisfatti e interessati alla lingua rispetto ad una lezione di tipo più tradizionale.

Queste tipologie di esercizi risultano utili specialmente per studenti universitari, per il grado di difficoltà e impegno che comportano; tuttavia, alcuni studi riportati a fine capitolo mostrano come i corpora possano essere adattati anche a livelli inferiori di educazione (Sealey & Thompson 2004; Geist & Hahn 2012).

Infine altri autori hanno investigato se gli studenti abbiano una preferenza e se ottengano prestazioni migliori per queste attività in forma stampata o elettronica (Chujo and Oghigian 2012, Boulton 2012): poiché le differenze rilevate negli studi risultano minime in entrambi i casi, e considerando che entrambe le modalità presentano sia vantaggi e svantaggi (l'utilizzo di un computer può essere infatti sia vantaggioso in termini di motivazione e coinvolgimento, che un impedimento se non se ne ha la disponibilità o se non si è particolarmente abili ad adoperarlo, mentre i file in forma stampata sono più facili e veloci da utilizzare ma permettono di usufruire di meno dati e non si discostano dalla forma tradizionale usata in classe), essi giungono alla conclusione che è auspicabile un

uso alternato dei due, in quanto possono essere utilizzati per tipi diversi di esercizi.

Successivamente, nel secondo capitolo mi soffermo solamente sul tipo di corpora utilizzati nella mia ricerca, ossia i *learner corpora*: come detto in precedenza, essi sono formati da collezioni di testi prodotti da studenti di una lingua (principalmente inglese, anche se negli ultimi anni si è assistito ad un numero crescente di corpora in altre lingue) di altra madrelingua.

Questo tipo di corpora può offrire un tipo diverso, ma egualmente valido, di aiuto per studenti ed insegnanti: infatti i testi prodotti da studenti di un'altra lingua possono essere comunque utilizzati in classe per approfondire argomenti o creare esercizi, ma a differenza dei *native corpora* non rappresentano la lingua autentica da emulare, ma mostrano come la lingua in questione viene prodotta quando studiata come seconda lingua o lingua straniera. Spesso questo si traduce in forme insolite di espressione o in veri e propri errori, che però si possono rivelare comunque utili per determinare le aree in cui gli studenti rilevano più fatica; per questo, nel condurre studi a partire da questi corpora, si analizzano, oltre ai testi, anche informazioni degli studenti stessi come il paese di provenienza, la totalità e il livello delle lingue parlate o eventuali periodi all'estero, che costituiscono importanti fattori che influenzano la loro produzione sia in positivo che in negativo. In particolare, gli studi tratti dai *learner corpora* possono essere di due tipi: il confronto può infatti essere fatto con *native corpora*, con lo scopo di individuare le principali differenze tra la produzione di madrelingua e non, oppure tra più *learner corpora*, al fine di individuare le caratteristiche peculiari di ogni lingua e rilevare eventuali similitudini o differenze in studenti della stessa lingua straniera.

Uno strumento che è particolarmente utile con questo tipo di corpora è quello dell'annotazione, che permette di modificare i testi, ad esempio assegnando le categorie di appartenenza di specifiche parole, e di poterle in seguito individuare con più facilità; questa modalità risulta infatti particolarmente efficace quando si considerano gli errori di un corpus, poiché annotarli permette di visualizzarli e successivamente categorizzarli in base alla tipologia in modo rapido ed eventualmente anche di poter inserire a seguito le relative forme corrette.

La metodologia del *data-driven learning* come accennato può essere utilizzata anche con questo tipo di corpora: anche esempi con forme inappropriate o errori possono infatti essere utilizzati nelle compilazioni dei dizionari, che da qualche decennio (il primo

pubblicato nel 1993) presentano infatti delle sezioni apposite dove gli errori più comuni sono trascritti seguiti dai corrispondenti usi corretti; sezioni simili si possono trovare anche in alcuni libri di testo, e possono essere sfruttate dagli insegnanti per rinforzare quegli aspetti della lingua che risultano più difficili per i loro studenti. Inoltre, durante le lezioni, analizzare esempi dove sono presenti degli errori, prodotti anche dagli studenti stessi, ed eventualmente compararli con altri tratti dai *native corpora* al fine di individuare le correzioni, permette agli studenti di avere un'ottica più positiva nei loro confronti e di non vederli più solo con paura o scoraggiamento.

Numerosi studi sono stati tratti da *learner corpora*: il principale per quanto riguarda la lingua inglese è l'ICLE (*International Corpus of Learner English*), fondato da Sylviane Granger, che comprende testi composti da studenti universitari provenienti da 16 nazionalità. Un numero così elevato di lingue permette di investigare numerosi aspetti rispetto a come la lingua inglese viene prodotta, di individuare parole o strutture in particolare che vengono usate troppo, troppo poco oppure in modo errato, e di poterle confrontare tra di loro: ad esempio, studi hanno considerato categorie grammaticali della lingua per rilevare se ne viene fatto un uso corretto o eccessivo, come nel caso di Granger & Paquot (2009) che hanno preso in considerazione la categoria dei verbi, oppure tramite la stessa procedura possono essere analizzate anche singole parole particolarmente frequenti e rilevanti appartenenti alla categoria in questione.

Nell'ultima parte del capitolo ho preso in considerazione studi di questo tipo riguardanti solamente studenti di lingua italiana e tedesca, in modo da poter avere già un'idea sui tratti peculiari della loro produzione in lingua inglese prima di affrontare la mia ricerca personale. Per quanto riguarda gli studenti italiani, gli studi riportati hanno mostrato come essi facciano particolare uso di forme contratte e domande dirette, che denota una trasposizione del linguaggio orale anche in quello scritto, di dislocazioni, probabilmente in quanto metodo principale studiato in classe per dare rilevanza ad un elemento in particolare all'interno della frase, e di connettori, che dimostra una preferenza per uno stile di scrittura più retorico. Inoltre, altre informazioni che è stato possibile notare sono una tendenza a scrivere frasi più lunghe rispetto ai nativi di lingua inglese e a produrre errori derivanti dall'influenza della lingua italiana, come ad esempio la forma plurale di alcuni sostantivi non numerabili come *information*. Per quanto riguarda gli studenti tedeschi, gli studi analizzati hanno riportato come essi abbiano una particolare tendenza

ad intensificare gli aggettivi e ad usare dislocazioni, allo stesso modo degli studenti italiani e principalmente per riferire opinioni soggettive, e modali, mentre mostrano più difficoltà con il tempo verbale del passato remoto, spesso confuso con l'imperfetto. Inoltre, uno studio condotto da Nesselhauf (2005) solo sulle *collocations* ha mostrato come gli studenti tedeschi ne facciano più uso rispetto a quelli nativi, producendo però anche qualche errore dovuto a traduzioni letterali dalla lingua tedesca, specialmente in quelle riguardanti verbi e sostantivi: per questo sarebbero consigliabili più esercizi mirati su questo tema, e l'ausilio di *concordances* tratte dai corpora in questo caso potrebbe essere particolarmente utile ricercando le due categorie grammaticali sovracitate e analizzandone i contesti.

Infine, uno sguardo alla forma orale della lingua attraverso il LINDSEI (*Louvain International Database of Spoken English Interlanguage*), il principale *spoken corpus* che riguarda l'inglese, ha mostrato attraverso altri studi come sia studenti italiani che tedeschi non siano particolarmente fluenti nella lingua: in particolare, quelli italiani fanno un uso minimo di avverbi o altri strumenti per creare un'omogeneità nel discorso (*discourse markers*), mentre quelli tedeschi tendono a produrre orazioni molto più brevi rispetto ai nativi e con pause più lunghe.

Prima di procedere con la mia personale ricerca su queste due categorie di studenti riguardo *get*, ho introdotto il verbo nel terzo capitolo, descrivendone i principali significati che può assumere e le forme in cui può essere trovato, in particolar modo in *phrasal verbs* e *idioms.* Essendo uno dei verbi più comuni della lingua inglese, esso può infatti essere considerato un verbo ad alta frequenza (*high-frequency verb*) nonché polisemico, in quanto presenta una notevole varietà di significati a seconda del contesto in cui è inserito: infatti se seguito da un sostantivo può assumere il significato attivo di "ottenere" (*get a job*) o passivo di "ricevere" (*get money*), oppure se seguito da un aggettivo può assumere o il senso incoativo, quando esprime un cambiamento di stato (*get bored*), o passivo. In particolare, il passivo con *get* rappresenta un'alternativa a quello più classico con il verbo essere *to be*, con la differenza che viene usato principalmente nel descrivere delle situazioni connotate negativamente, come negli esempi *get killed*, *get caught* o *get arrested*, che al contrario delle relative forme con *to be*, accentuano il cambiamento di stato causato dall'azione e danno maggiore attenzione al soggetto rispetto all'agente della frase, caratteristica confermata dal fatto che spesso l'agente

introdotto dalla preposizione *by* non è affatto presente. Inoltre *get* può anche esprimere movimento (spesso quando seguito dalla preposizione *to,* come in *get to school*), causa, che viene espressa da *get* seguito da un argomento che può venire espresso da più strutture (infinito, verbo alla forma -ing, un aggettivo, ecc, come in *get someone to do something*), l'inizio di uno stato, il cui uso viene denominato ingressivo (*get to understand a situation*), o avere un significato metaforico, come sinonimo di "capire" (*nobody gets me*). Nella sua forma al passato *got*, il verbo può anche assumere il significato ausiliare di "avere" nella forma *have got* e quello modale di "dovere" nella forma *have got to*.

Per quanto riguarda *phrasal verbs* e *idioms*, i primi si riferiscono a dei verbi formati da due o più elementi, dove uno è il verbo principale, mentre gli altri possono essere preposizioni o avverbi, denominati particelle (es. *get down*, *get around*); i secondi invece hanno una struttura più varia, in quanto possono anche presentare sostantivi al loro interno (es. *get the wrong end of the stick*). Ciò che hanno in comune è che hanno un significato che non è deducibile a partire dalle singole componenti che li formano, a cui è dovuta anche la regola per cui non è possibile inserire tra di esse altri elementi della frase: per queste ragioni costituiscono un argomento particolarmente difficile per gli studenti.

Gli studi riportati fanno riferimento inizialmente ai cosiddetti *high-frequency verbs* che, nonostante siano tra i primi che gli studenti incontrano nei loro percorsi di studi, risultano difficili da imparare nella loro interezza e costituiscono una delle aree più problematiche anche per studenti di livello più avanzato. Questi verbi vengono infatti usati principalmente con il loro significato principale, e ciò che risulta più complicato sono i significati secondari, che spesso coincidono con forme quali le *collocations*: un motivo può essere identificato nello studio condotto da Gouverneur (2008), che ha mostrato come infatti essi siano spesso posti in secondo piano nei libri di testo, in quanto spesso menzionati ma non con la giusta attenzione. Gli studi analizzati hanno riportato risultati contrastanti, in quanto secondo alcuni molti di questi verbi risultano notevolmente meno frequenti nei *learner corpora* rispetto che ai *native corpora* e, quando prodotti, spesso usati in modo improprio, in quanto altri verbi risulterebbero più adatti al contesto (Gerckens & Gans 2015), mentre altri, per lo stesso motivo, hanno mostrato come risultino molto più usati dagli studenti che dai nativi (Ringbom 1998): essendo i primi verbi studiati e più frequentemente incontrati della lingua, spesso vengono infatti utilizzati in quanto considerati "sicuri", mentre per alcune espressioni specifiche risultano

errati. Sarebbe perciò utile condurre più ricerche di questo tipo, in modo da poter determinare se i verbi specifici e le nazionalità degli studenti sono gli unici fattori essenziali nel determinarne i risultati.

Per quanto riguarda i *phrasal verbs*, che in molti casi presentano un *high-frequency verb* nella loro costruzione, risultano anch'essi difficili da imparare e produrre dagli studenti, in particolar modo per la loro struttura, per la loro vasta quantità e per il fatto che spesso hanno più di un solo significato. Infatti, da alcuni studi è emerso come nelle produzioni di *learner corpora* siano presenti molti meno *phrasal verbs* che in quelle dei *native corpora*, e di come spesso gli studenti, sia coloro familiari con questa struttura dalla loro madrelingua che non, preferiscano usare verbi singoli con gli stessi significati come alternative (es. Chen 2013). Alcuni autori hanno suggerito come possibile metodo di studio partire dalle particelle più frequenti e osservare le combinazioni principali che formano con i verbi e viceversa, esercizio che può essere notevolmente facilitato dall'uso delle *concordances* sovra menzionate. Inoltre, feedback ottenuti dagli studenti stessi che hanno partecipato agli studi hanno mostrato come leggere e guardare film in lingua inglese abbiano un effetto positivo in quanto facilitano ricordo e comprensione e migliorano il rendimento di questi verbi, mentre lo stesso non è manifestato da attività come ascoltare musica o navigare nei social network: questo è stato il risultato che mi ha colpito maggiormente, in quanto personalmente ho sempre trovato l'attività di ascoltare musica molto più utile della lettura, in quanto permette di sentire con una frequenza più elevata le parole e di poterle perciò ricordare con più facilità.

Infine, nell'ultima parte del capitolo avevo intenzione di fornire studi effettuati in passato sul verbo stesso, in modo da poter poi successivamente confrontarli con il mio: tuttavia, ho avuto accesso ad un solo studio riguardante *get*, precisamente dell'autore Coto-Villalibre (2016) che ha considerato la componente di Hong Kong dell'ICLE. In particolare, l'autore prende in considerazione solo le costruzioni participiali del verbo e rileva come tra di esse le costruzioni con aggettivi siano le più frequenti, seguite da quelle passive e infine da quelle idiomatiche, mentre un confronto finale con la componente indiana dello stesso corpus mostra come questi ultimi facciano un uso più rilevante di *get* attraverso i passivi. Questo studio, che in ogni caso non fa riferimento ad una panoramica completa degli usi del verbo, mostra perciò come la madrelingua degli studenti abbia un'influenza particolare nella produzione del verbo e nella frequenza con cui è utilizzato

nei suoi vari significati.

Il quarto ed ultimo capitolo è, come accennato in precedenza, dedicato al mio personale studio su come studenti italiani e tedeschi usano il verbo *get*: per i dati ho usufruito dei testi presenti nelle componenti italiana e tedesca dell'ICLE e come software per ricavare tutti gli esempi dove viene prodotto il verbo ho utilizzato *AntConc*. In particolare, con la dicitura "get*" ho ricavato gli esempi in cui il verbo è espresso alla forma presente, mentre "got" per i casi in cui è espresso alla forma passata.

Un primo dato da sottolineare è che entrambi i sub-corpora sono tra i più grandi dell'ICLE, in quanto presentano un numero elevato di parole: nonostante ciò, quello tedesco si discosta in maniera notevole da quello italiano nella produzione di *get*, poiché presenta molti più esempi del verbo (597 contro i 198 italiani).

Analizzando in primo luogo gli studenti italiani, è stato possibile notare non solo come non producano molti esempi del verbo, ma anche come, in riferimento ai significati elencati in precedenza, non mostrino di conoscerli tutti. Infatti, è possibile trovare *get* seguito da un sostantivo, che rappresenta la modalità più frequente nel sub-corpus (87 esempi), sia in forma attiva nel senso di 'possedere' che passiva nel senso di 'ricevere'; il secondo uso più frequente di *get* è quando seguito da un aggettivo (74 esempi), sia nella forma incoativa che in quella passiva; il verbo è presente anche sotto forma di *phrasal verb* o *idiom*, per un totale di 26 esempi, mentre i significati di movimento e di causa sono espressi rispettivamente solamente in uno e in due esempi. L'uso ingressivo che denota l'inizio di uno stato risulta invece completamente assente.

Prendendo in considerazione gli studenti tedeschi, essi mostrano diversamente un ampio uso di tutti i significati menzionati. Anche nel loro caso, *get* seguito da un sostantivo rappresenta la forma più comune, con un totale di 188 esempi; 147 sono invece gli esempi del verbo seguito da aggettivo, che non rappresentano tuttavia la seconda categoria più frequente come nel caso del corpus italiano, in quanto gli studenti tedeschi producono una quantità più vasta di *phrasal verbs* e *idioms*, per un totale di 168 esempi. Infine, anche il significato di movimento è espresso molto più frequentemente (51 esempi), così come quello di causa (25 esempi), mentre l'uso ingressivo, benché sia presente, è prodotto solamente in quattro esempi.

È possibile notare dunque come nella produzione degli studenti tedeschi il verbo *get* sia molto più presente in molti significati diversi, tanto che la totalità di esempi di *get* +

sostantivo è quasi equivalente alla totalità di *get* in generale degli studenti italiani; altre differenze possono essere rilevate nella forma *get* + aggettivo, in quanto gli studenti italiani producono più frasi passive mentre quelli tedeschi più frasi incoative.

Per quanto riguarda gli errori presenti nei due sub-corpora, in entrambi se ne possono individuare alcuni dovuti all'influenza della madrelingua degli studenti, in quanto vengono usati i due principali verbi equivalenti di *get*, *ottenere* e *bekommen*, anche in casi dove non sono appropriati; inoltre, un errore comune alle due categorie è la forma *get used to*, poiché espressa da entrambe erroneamente quando fatta seguire dalla forma all'infinito e non dalla forma in -ing.

La parte finale del mio studio è dedicata al corpus LOCNESS, formato sia da testi prodotti da studenti britannici che americani: uno sguardo agli esempi di *get* presenti in esso ha mostrato come gli americani ne facciano più uso rispetto ai britannici, e un confronto con le due categorie di studenti dell'ICLE mostra come questi ultimi presentino più esempi per ogni categoria, confermando il risultato che spesso nei *learner corpora* sono presenti più *high-frequency verbs* che nei *native corpora*, a sola eccezione del significato di causa, molto più utilizzato dai nativi. Un confronto individuale tra le quattro categorie di studenti mostra tuttavia come la differenza sia principalmente dovuta agli studenti tedeschi, che producono il verbo non solo più degli studenti italiani, ma anche di quelli britannici e americani, con la sola eccezione della forma passiva con una differenza tuttavia minima. Un ultimo confronto con i risultati dello studio di Coto-Villalibre (2016) ha evidenziato delle similitudini tra gli studenti italiani e tedeschi e quelli di Hong Kong, ossia come, nelle forme *get* + aggettivo, gli usi passivi ed incoativi vengano utilizzati con maggior frequenza rispetto a *phrasal verbs* e *idioms*.

In conclusione, questo studio ha mostrato come gli studenti tedeschi abbiano un'ampia conoscenza del verbo *get* e dei suoi significati, mentre quelli italiani mostrano molte più difficoltà: in particolare, l'area dei *phrasal verbs* è quella che ha mostrato risultati più contrastanti, in quanto gli studenti tedeschi ne producono in quantità elevata e attraverso più significati commettendo tuttavia degli errori, mentre la produzione degli studenti italiani è al contrario limitata. Per entrambe le categorie sarebbero perciò utili esercizi tratti da *concordance lines* di corpora, per i primi in particolare per fare pratica su come sono strutturati, in quanto gli errori sono soprattutto grammaticali e sintattici, e per i secondi sui loro significati, in quanto tendono a mostrare meno padronanza nel loro uso.

Inoltre, questa ricerca è limitata da più punti di vista: dal momento che prende in considerazione un solo verbo, un solo corpus e due sole categorie di studenti della lingua, sarebbe interessante scoprire tramite ulteriori studi se questi risultati vengono confermati su altri verbi ad alta frequenza simili a *get* come *make* o *take,* su altri corpora dove sono presenti le componenti italiana e tedesca (come ad esempio uno *spoken corpus* come il LINDSEI) e su altre categorie di studenti, per poter effettuare un confronto in modo più completo.