# Second language speakers' use of multi-word verbs in spoken communication: Evidence from the Trinity Lancaster Corpus

Irene Marín Cervantes

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Linguistics and English Language

Lancaster University

June 2019

**Declaration**

I declare that this thesis represents the outcome of my own original research and that it has not been submitted in substantially the same form for the award of a higher degree in this University or elsewhere.

Irene Marín Cervantes Signed ……

**Acknowledgements**

I would first like to thank Dr. Dana Gablasova, my main supervisor, for her professional guidance, insightful comments and suggestions which helped me find direction when I saw none. Her encouragement throughout the process of writing this thesis and her advice (PhD-related and otherwise) have been invaluable. I am also deeply grateful to Professor Tony McEnery, my co-supervisor, for his suggestions and comments particularly on the results chapter of this thesis. Special thanks also go to Dr. Vaclav Brezina for his willingness to share his vast knowledge on corpus statistics with me and for taking the time to answer my questions about this topic.

This thesis would not have been possible without the scholarship I was awarded by the University of Costa Rica. I am particularly indebted to the Department of Modern Languages for supporting my application to study in Lancaster University.

My family in Costa Rica merit special thanks: María Elena (my mother), Arturo (my father) and María Antonieta (my sister). Their unconditional love, trust, patience, and much needed calls have been essential during my time in Lancaster.

I am also grateful to Mayra Solís, whose wise words always came in the right place and time, for encouraging me to pursue a PhD.

Finally, I would like to express my heartfelt gratitude to my daughter, Isabel, and my husband, Carlos. This thesis is dedicated to them. Having both of them by my side has been my main source of strength during the course of my PhD. I thank Isabel for making

me a mother in the middle of my PhD journey. Her love, hugs and laughs every single day since she was born have made any sacrifice to finish this thesis worth it. I thank Carlos for showing me his love through music in all possible ways and for being an unending source of support. I am also grateful to him for putting up with me during countless conversations about research and corpus linguistics and for expressing genuine interest in my PhD project.

**Abstract**

The present study investigates L2 speakers' use of multi-word verbs (MWVs) in spoken, interactive communication. It focuses on two specific types of MWVs: phrasal verbs (e.g. *carry out, get up*) and phrasal prepositional verbs (e.g. *look up to, get along with*). Its main aim is to use corpus methods in order to examine i) how L2 speakers use MWVs in terms of frequency, lexical verb and particle productivity, number and type of non-canonical MWV forms as well as polysemy, and ii) whether learner variables (i.e. L2 proficiency and L1 background) and situational variables (i.e. task type) mediate L2 speakers' use of MWVs.

The study draws on data from the Trinity Lancaster Corpus (Gablasova, Brezina & McEnery, forthcoming), a large-scale POS-tagged corpus of L2 spoken production consisting of 4.2 million running words. More specifically, the speech of 1,348 L2 speakers was analysed. The speakers were at three different levels of L2 proficiency (i.e. B1, B2 and C1-C2) according to the Common European Framework of Reference and came from three different L1 backgrounds: Chinese, Italian and Spanish. MWVs were automatically extracted via the Sketch Engine (Kilgarriff et al., 2014) using tags based on Corpus Query Language (CQL). The automatic extraction was followed by a manual scrutiny of the results to discard false positives. For the analysis of data, both descriptive and inferential statistics were used.

With respect to overall MWV frequency patterns, the results indicated that there was a relatively low MWV representation in L2 production, with a few high-frequency MWVs accounting for the majority of MWV occurrences in the corpus. A variety of

non-canonical MWVs were also found; most of these forms were one-off occurrences produced mainly by the most advanced L2 speakers. Four patterns of non-canonical MWV use were identified including, for example, the combination of a lexical verb with a redundant particle and the creation of novel MWVs. The study also provided evidence of a frequency-polysemy phenomenon in which the most polysemous MWVs that the L2 speakers produced featured among the most frequent MWVs in the corpus.

In terms of the relationship between learner variables and L2 MWV production, the results showed that the main effect for L2 proficiency on MWV use was not statistically significant, which indicated that gains in proficiency did not seem to translate into a high MWV frequency. In contrast, L1 background was found to be a significant factor mediating MWV use. From the three L1 backgrounds studied, Chinese L1 speakers appeared to use more MWVs per thousand words than Italian and Spanish L1 speakers. Moreover, the results revealed a large inter-speaker variation across proficiency levels and L1 backgrounds, and they also provided evidence of topic effects on MWV use given that particular MWVs recurred in the context of specific topics.

Regarding task-related variables, the effect of speaking task type on MWV production was found to be statistically significant. The only monologic, pre-planned task contained not only a higher number but also a wider range of MWVs than the dialogic, unplanned tasks that were analysed. The MWVs in the monologic task tended to function as part of transitions in the context of delivering oral presentations (e.g. *now let's move on to the next point*). The same MWVs found in the monologic task also occurred in the dialogic tasks. However, rather than serving as part of transitions in those tasks, these MWVs expressed a variety of meaning senses.

The findings of this study both complement and add to previous research on MWVs in L2 contexts, particularly by showing important trends of MWV use in a spoken, interactive context and across a variety of L2 proficiency levels, L1 backgrounds and speaking tasks. The study also contributes to broaden our understanding of the MWV knowledge that L2 users possess, which has been found to enhance L2 speakers' communicative competence and facilitate both language production and comprehension.

**Table of Contents**

**List of tables**

**List of figures**

**List of abbreviations**

| | |
|---|---|
| ANOVA | Analysis of Variance |
| BNC | British National Corpus |
| BNC2014 | British National Corpus 2014 |
| CASS | Centre for Corpus Approaches to Social Science |
| CEA | Computer-aided Error Analysis |
| CEECS | Corpus of Early English Correspondence Sampler |
| CEFR | Common European Framework of Reference |
| CIA | Contrastive Interlanguage Analysis |
| COCA | Corpus of Contemporary American English |
| CQL | Corpus Query Language |
| GESE | Graded Examinations in Spoken English |
| GloWbe UK | Corpus of Global Web-based English UK |
| GloWbe US | Corpus of Global Web-based English US |
| ICE | International Corpus of English |
| ICLE | International Corpus of Learner English |
| L1 | First language |
| L2 | Second language |
| LINDSEI | Louvain International Database of Spoken English Interlanguage |
| LOCNEC | Louvain Corpus of Native English Conversation |
| LOCNESS | Louvain Corpus of Native English Essays |
| LSWE | Longman Spoken and Written English Corpus |
| MICASE | Michigan Corpus of Academic Spoken English |
| MWU | Multi-word unit |
| MWV | Multi-word verb |
| POS | Part of speech |
| PPCMBE | Penn Parsed Corpus of Modern British English |
| PPV | Phrasal prepositional verb |
| PV | Phrasal verb |
| RQ | Research question |
| SLA | Second language acquisition |
| TLC | Trinity Lancaster Corpus |
| TLC-L1 | Trinity Lancaster Corpus for L1 English Interaction |

VOICE Vienna-Oxford International Corpus of English

# Chapter 1: Introduction

## 1.1. Introduction

The present study looks at L2 speakers' use of multi-word verbs (hereafter MWVs) in L2 spoken, interactive communication. Two specific types of MWVs are analysed: phrasal verbs (henceforth PVs) (e.g. *pick up, take off*) and phrasal prepositional verbs (henceforth PPVs) (e.g. *get along with, look forward to*). Unless otherwise specified, the term MWVs refers to those two types of verbs only.

The aim of this chapter is twofold. First, it provides an overview of the theoretical background of the present study. Second, it describes its research aims and significance. Section 1.2 sets out to define what multi-word units are and highlights their importance in the contexts of L2 production and comprehension. Next, section 1.3 is concerned with the two main corpus-based approaches to the study of learner language (i.e. Computer-aided Error Analysis and the comparative research design), their characteristics and the criticism they have received. Section 1.4 outlines the reasons that motivated the study. In section 1.5, the aims and scope of the thesis are reported, followed by a description of the theoretical and methodological contributions of the study in section 1.6. Finally, section 1.7 presents an outline of this thesis.

## 1.2. The importance of multi-word units in learner language

Multi-word units[1] (MWUs) are broadly defined as lexical combinations, rather than single words, which appear to be stored and retrieved as wholes from memory (Wray, 2002; Grant & Bauer, 2004). Because of their pervasiveness in English, MWUs have been increasingly attracting a great deal of scholarly attention (e.g. Wray, 2002, 2008, 2019; Sinclair, 2004; Meunier & Granger, 2008; Granger, Paquot & Rayson, 2009; Ellis, 2012; Wood, 2015; Siyanova-Chanturia & Pellicer-Sánchez, 2019). An important part of the literature on MWUs has been concerned with defining and classifying the different types of MWUs as well as describing the criteria that can be used to identify them. The criteria that are often mentioned include aspects like adjacency (i.e. whether the components of a MWU can occur next to each other or be separated), grammatical fixedness, grammatical completeness, lexical variability (i.e. whether the components of the MWU can be replaced by others with similar meaning), length, degree of compositionality, institutionalization (i.e. whether the MWU is conventionalised in the language community), frequency of occurrence, and semantic prosody (i.e. whether the MWU connotes a particular attitudinal meaning) (Schmitt & Carter, 2004; Nation & Webb, 2011; Columbus, 2013).

As evidenced in the abundant research on the topic, MWUs are a fundamental component of both written and spoken communication. They have been found to comprise a substantial part of the language that is produced across contexts and language domains (Szudarski, 2018). In fact, Altenberg (1998) highlights the high

[1] In this thesis, the terms multi-word units and formulaic units will be treated as synonyms and will be used interchangeably.

frequency of MWU occurrence by noting that MWUs account for 80 percent of the spoken English language. From the perspective of second language acquisition and use, the fact that MWUs are ubiquitous in English implies that L2 speakers will encounter them often and that they will have to use those units appropriately in order to communicate effectively in the L2 and eventually reach a higher level of L2 proficiency (Crossley, Salsbury & McNamara, 2015; Siyanova-Chanturia & Pellicer-Sánchez, 2019). The associated benefits of using MWUs for L2 speakers tend to relate to improvements in four main areas: fluency, idiomaticity, accuracy, and pragmatic competence (Siyanova-Chanturia & Nation, 2017). With respect to the first of these areas, it has been pointed out that knowledge of MWUs facilitates fluent speech. The relationship between fluency and MWU knowledge is based on observations regarding the way in which MWUs are stored and retrieved as wholes or 'chunks' from memory. Despite the fact that further research is needed to explore the mechanisms underlying the storage and retrieval of MWUs (Siyanova-Chanturia, 2015), they have been shown to enjoy a processing advantage given that language users can retrieve them as 'ready-made' items (Conklin & Schmitt, 2008; Nation & Webb, 2011; Schmitt & Redwood, 2011; Götz, 2013; Dóczi & Kormos, 2016). This means that speakers do not have to spend time generating word-by-word phrases, a process that often results in more pauses and hesitations (Skehan, 1998; Boers et al., 2006) and therefore slows L2 production.

Second, mastery of MWUs tends to translate into a more idiomatic language use, which in turn helps L2 speakers appear more native-like. MWUs are considered idiomatic (Wray, 2002; Boers et al., 2006). Idiomaticity in this context refers to the ways in which "the members of the relevant speech community" (Wray, 2002, p. 20) would express

themselves. L2 speakers can thus benefit greatly from knowing and using MWUs, especially in situations where their speech is grammatical but may not sound 'natural' (Pawley & Syder, 1983; Conklin & Schmitt, 2008; Wood, 2015) or reflect the way in which L1 speakers would convey the same message.

The use of MWUs has also been found to result in greater linguistic accuracy. According to Boers et al. (2006), MWUs that are correctly committed to memory represent "zones of safety" (p.247) and can provide L2 speakers with opportunities to produce error-free speech. Because a large portion of the language in a MWU, if not the whole unit, is ready-made and available to the L2 speaker, there is a lower risk of making mistakes trying to select and put together the right words.

In addition to gains in fluency, idiomaticity and linguistic accuracy, L2 speakers' pragmatic competence can greatly benefit from their use of MWUs. It is well established that MWUs have a prominent role in pragmatic language use (Wood, 2002; Bardovi-Harlig, 2012, 2013, 2019; Siyanova-Chanturia & Pellicer-Sánchez, 2019). They can realise different functions in communication (e.g. expressing agreement or disagreement, thanking, expressing lack of understanding, and requesting help), and thus be of great value to L2 speakers in situations where the communicative context demands language to realise those functions. Thus, the pragmatic utility of MWUs relates to the fact that they can help L2 speakers achieve different communicative goals, which in turn is likely to result in successful communication.

In summary, the importance of MWUs in learner language lies in their practical value. They contribute to the development of L2 speakers' proficiency and may help them

produce more fluent, idiomatic and accurate language in pragmatically appropriate ways. It is also worth pointing out that the knowledge of MWUs can be highly motivating for L2 speakers (Wray, 2002) and even increase their willingness to communicate and their ability to express ideas.

## 1.3. Corpus-based approaches to the study of learner language

The analysis of L2 speaker language through the lens of corpus linguistics has typically involved using one of the following two approaches: Computer-aided Error Analysis and the comparative research design. In what follows, the origin and main characteristics of both approaches will be described. Attention will also be given to arguments questioning the appropriateness of the two approaches.

### 1.3.1. Computer-aided Error Analysis

Computer-aided Error Analysis (CEA) (Dagneaux, Denness & Granger, 1998) is an approach to learner language that consists of the analysis of L2 speakers' errors which have been annotated in a learner corpus (Granger, 2003). The theoretical basis of CEA is that of traditional error analysis (e.g. Corder, 1971, 1981), a methodology whose main purpose is to describe L2 speakers' language by locating and analysing errors in their production (James, 1998).

The literature on CEA has emphasised the fact that it is a descriptive approach. This means that the CEA helps to identify errors found in L2 speakers' production, but it does not attempt to provide explanations for the occurrence of those errors (Dagneaux, Denness & Granger, 1998). While the comparative research design has been considered

useful in identifying cases where linguistic features occur more or less frequently in two or more corpora (see section 1.3.2), the CEA appears to be more suitable for investigating instances of 'misuse' (Granger, 2015). It is also worth pointing out, however, that the CEA has been described as an approach that may not only focus on identifying errors, but it can also be used in examining instances of correct language use (Granger, 2009).

In order to identify errors, the CEA first requires the development of an error taxonomy which is later used in an error-tagging system to assign a tag to each error in the corpus (Díaz-Negrillo & Fernández-Domínguez, 2006; Reznicek, Lüdeling & Hirschmann, 2013; Lüdeling & Hirschmann, 2015). The level of specificity of the error tags may vary. Often error tags indicate the general error category (e.g. grammatical) and are followed by sub-codes specifying the type of error (e.g. tense, auxiliary or voice) (Dagneaux, Denness & Granger, 1998). Moreover, the error-tagging systems used in the CEA tend to be rather flexible because researchers can often add or change error tags according to their research needs.

As acknowledged in the literature on the CEA, one of the main advantages of this type of approach is that it allows automatic error searches, counts, and analyses of learner language which were not possible in the past (Granger, 2003; Díaz-Negrillo & Fernández-Domínguez, 2006; Lüdeling, Hirschmann & Shadrova, 2017). Furthermore, the CEA can provide insights into L2 speakers' ideas and hypotheses about certain linguistic concepts (Lüdeling & Hirschmann, 2015). Despite these advantages, the CEA has attracted criticism on the grounds that it still appears to suffer from some of the same limitations of the traditional error analysis. A particular unresolved issue in the

CEA relates to the high level of subjectivity involved in the development and assignment of error categories given that it is sometimes possible to assign more than one tag to the same error (Granger, 2009; Callies, 2015).

### 1.3.2. Comparative research design

Comparative methodologies have been a common approach to the study of learner language even before the advent of learner corpora. For example, contrastive analyses where L2 speakers' L1 and L2 were systematically compared were embraced by many scholars during the 1960s and 1970s (Odlin, 1989; Selinker, 1992). In the context of learner corpora, the comparative research design has been employed in two specific types of comparison: i) a comparison between a native speaker language variety (L1) and a non-native speaker language variety (L2), and ii) a comparison between two or more non-native speaker language varieties or interlanguages (L2s) (Granger, 1996, 1998a). Both types of comparisons form the basis of the Contrastive Interlanguage Analysis (CIA) framework developed by Granger (1996). In order to conduct the first type of comparison (i.e. L1-L2), the CIA requires a control L1 corpus which serves as the reference against which L2 data is compared. For L2-L2 comparisons, learner corpora or sub-corpora that differ in terms of learner variables (e.g. age, mother tongue, L2 proficiency level, etc.) are needed. The CIA is certainly influenced by the contrastive analyses performed more than thirty years ago. However, unlike the traditional contrastive approach, the CIA can also be employed to compare the language produced by different groups of L2 speakers (i.e. different varieties of the same language) without the use of any L1 variety as reference (Hasselgård & Johansson, 2011).

In a more recent version of the CIA (Granger, 2015), the term 'native speaker language variety' was replaced by 'reference variety', a more encompassing term highlighting that different language varieties (and not only native speaker language) can be used for comparison purposes. Along the same line, the term 'interlanguage varieties' was adopted instead of 'non-native speaker language varieties' to draw attention to the various possible variables that can be considered in L2-L2 comparisons (e.g. proficiency level, age, L1 background, task types, etc.). In addition, the updated CIA model expands L1-L2 comparisons in the sense that it allows the comparison of L2 speaker data against one or more L1 corpora.

The L1-L2 and L2-L2 comparisons provide different information about patterns of L2 use. On the one hand, L1-L2 comparisons have been found to be particularly useful to identify linguistic features that are more or less frequently used by L2 speakers than by L1 speakers in similar situations. In other words, they can help study instances of 'overuse' and 'underuse' in L2 production relative to L1 production (Granger, 1998a, 2009). On the other hand, the main purpose of L2-L2 comparisons is to identify and differentiate patterns of L2 use that are characteristic of a specific L2 group (e.g. patterns that occur in the production of speakers of a particular L1 or at a specific proficiency level) from those that are general and can be found in the production of L2 speakers irrespective of their L1 or proficiency level (Granger, 1996; Callies, 2015). Both types of comparison need not be mutually exclusive and can indeed be integrated.

The rationale behind the comparison of (inter)language varieties relates to the fact that analysing and interpreting information about only one type of language variety may not be informative enough and is likely to result in many patterns of L2 use going unnoticed

(Granger, 1998a, 2015). Also, since learner corpora (the same as general corpora) provide information about frequency, which would be difficult to interpret in isolation and without a point of comparison (Gablasova, Brezina & McEnery, 2017), they naturally lend themselves to a comparative approach.

Despite its extensive use in learner corpus studies, the CIA framework is not without limitations. Various issues have been raised regarding the application of a contrastive approach to the analysis of L2 speaker language. More specifically, the comparisons between L2 and L1 speaker language have drawn criticism. Three main arguments have been advanced against the use of this type of comparison. First, L1-L2 comparisons may fail to recognise L2 varieties as languages in their own right. Second, they appear to use the L1 speaker data as a yardstick against which L2 speaker production is measured, which is problematic given that there are other target language varieties that can serve as reference. Third, interpreting differences between L1 and L2 speaker data in terms of over- and underuse may imply that L2 speaker language is deficient because it does not conform to L1 speaker norms (Larsen-Freeman, 2014; Gablasova, Brezina & McEnery, 2017). Most of these arguments related to the comparison of L1 and L2 speakers' production have been countered (e.g. Granger, 2009) and others addressed when the CIA model was revised (Granger, 2015).

The CIA as well as the CEA represent useful ways to approach the study of naturally occurring L2 speaker data. However, since the focus of the present study is the description and analysis of MWV use by L2 speakers rather than a systematic analysis of MWV-related errors, the CEA was not adopted. Instead, the research design of this study is more in line with the comparative approach found at the core of the CIA. In

particular, the study follows an L2-L2 comparative analysis and does not draw on an L1 reference corpus (see section 3.2).

### 1.4. Motivation for the study

Vocabulary is centre stage in the process of learning a second language. It allows one to express ideas, understand those of others, describe objects and events, and convey meaning. The lack of L2 vocabulary can seriously hinder one's ability to communicate and comprehend language (Folse, 2004), which highlights the importance of vocabulary to master a second language. The general underlying motivation for this study is, therefore, to contribute to advancing our understanding of L2 vocabulary use by looking at a particular lexical feature: MWVs.

MWVs are a frequent and important component of formulaic language. In addition to facilitating language comprehension and production through improvements in fluency, accuracy and pragmatic competence (see section 1.2), MWV use can help speakers increase their range of expression in the L2, enhance their communicative competence and contribute to achieve high levels of L2 proficiency in both speech and writing (Bolinger, 1971; Cowie, 1993). Because of all these benefits associated with their use, MWVs represent a lexical feature worthy of attention.

The specific motivations for the study of MWVs are theoretical and pedagogical. With respect to the former ones, this investigation seeks to contribute to our theoretical understanding of MWV use in two ways. First, most corpus-based work on MWVs in L2 contexts has been based on written discourse (see section 2.3.2.3). While insights from written learner corpora are valuable without a doubt, there is a need to study how

L2 speakers use MWVs in spoken communication and to analyse the extent to which the patterns of MWV use in speaking differ from those observed in writing. A further theoretical consideration that motivated this study relates to the lack of evidence from corpus-based research about the relationship between i) L2 proficiency and MWV frequency and ii) task type and MWV frequency. At present, such evidence mainly comes from elicitation studies that have assessed intermediate and advanced L2 speakers' production of a limited set of MWVs in controlled tasks (see section 2.5). However, learner corpora have not been systematically exploited for these purposes.

Pedagogically, the study is motivated by a need to expand our knowledge of a verb type that has been considered one of the most challenging aspects in vocabulary teaching and learning (Sinclair, 1991; Cowie, 1993; González, 2010; Ke, 2017). Language practitioners often face the challenge of selecting the best methods to approach the teaching of MWVs and may be overwhelmed by the sheer quantity of MWVs that can be taught (Schmitt & Redwood, 2011; Garnier & Schmitt, 2016). It is also well known that MWVs pose great difficulty to L2 speakers (Sinclair, 2004; Siyanova & Schmitt, 2007; Waibel, 2008), and the reasons for this are manifold: the way they are represented in written form (two orthographic words as opposed to most one-word verbs), their apparent idiosyncratic meanings, their lack of correspondence with verbs in different L1s, their syntactic flexibility, and polysemous behaviour. In order for L2 speakers to overcome those difficulties in acquiring MWVs, it is necessary to gain a thorough understanding of L2 speakers' use of MWVs. This study is conducted in the hope that its findings can help contribute to this understanding.

## 1.5. Aims and scope of the thesis

The main aim of the present study is to provide insights into L2 speakers' use of MWVs in spoken, interactive communication and the individual and external factors that may play a role in this use. Two types of MWVs are examined: PVs (e.g. *carry out*) and PPVs (e.g. *look forward to*). In particular, the study analyses MWV production by L2 speakers from three proficiency levels according to the Common European Framework of Reference (CEFR) (i.e. B1, B2 and C1-C2) and from three L1 backgrounds (i.e. Chinese, Italian and Spanish) in different speaking tasks. The study also seeks to make use of a number of corpus methods in order to analyse MWV production. It draws on data from the Trinity Lancaster Corpus (Gablasova, Brezina & McEnery, forthcoming), a new and large corpus of L1-L2 spoken communication (see section 3.4). As opposed to the majority of corpus-based studies on MWV use in L2 contexts (see section 2.3.2.1), the present investigation does not compare L1 and L2 English varieties. Instead, it compares MWV production in L2 speaker groups that differ in terms of L2 proficiency and L1 background.

## 1.6. Significance of the study

The significance of this investigation is both theoretical and methodological. On the theoretical level, the study provides a comprehensive corpus-based description of L2 speakers' MWV use in spoken, interactive communication, an underexplored register in previous elicitation and corpus-based research on MWVs. Moreover, the study responds to recent calls for analysing L2 speaker language as elicited in various tasks (Tracy-Ventura & Myles, 2015; Alexopoulou et al., 2017; Caines & Buttery, 2018). To the best of my knowledge, no single corpus-based study has been conducted on the

effect of the type of speaking task on L2 speakers' MWV use. Therefore, this study goes one step further in the analysis of MWVs as it looks at their production in different monologic and dialogic tasks (see section 3.5.3). In terms of the effect of L2 proficiency on L2 use, the study adds to findings from earlier corpus-based research by reporting patterns of MWV use not only in advanced L2 production as it has been done previously (e.g. Waibel, 2008; Weirszycka, 2013; Gilquin, 2015b) but also in the speech of L2 learners at lower levels of proficiency. Also, while the link between L2 proficiency and MWV use has been explored mainly by means of elicitation studies (e.g. Hulstijn & Marchena, 1989; Liao & Fukuya, 2004) (see section 1.4), little corpus-based work has been done in this respect. The present study contributes to filling the gap in knowledge about the effect of L2 proficiency on MWV production by using learner corpus data in order to analyse whether the frequency and range of MWVs vary across a range of proficiency levels.

On the methodological level, the study goes beyond what has been done in previous research on MWVs in terms of three main aspects: corpus size, analysis of L1 background influence on MWV use, and manual coding of MWVs. First, this study draws on the largest corpus of L2 spoken, interactive English that is available to date, i.e. the Trinity Lancaster Corpus. The findings reported come from more than a thousand L2 speakers, which provides robust evidence for the trends observed. Second, this study extends previous corpus-based work on the effect of crosslinguistic differences on L2 MWV production not only by acknowledging the presence of MWVs in Chinese, Italian and Spanish (i.e. the three L1 backgrounds included in this study) but also by describing MWVs and similar verbal forms in those L1 backgrounds. This approach to the analysis of L1 background effects on MWV production was followed

i) to address methodological concerns raised in previous research on L1 influence (e.g. Granger, 1998a; Jarvis, 2000; Odlin, 2003; Alonso-Alonso, Cadiermo & Jarvis, 2016; Ringbom, 2016), and ii) to observe whether there are any lexical, syntactic or semantic differences between MWVs in speakers' L1s and in English, which may help to better understand the relationship between L1 background and speakers' MWV use. While descriptions of the target linguistic features as found in speakers' L1s have certainly been included in L2 corpus-based studies, they have not been provided for MWVs. Lastly, similar to earlier studies on the topic of MWVs, this study recognises the importance of manually coding MWVs after their automatic extraction given that POS-tagging errors are likely to occur (see section 3.10.2). However, not only is the manual coding process mentioned in this study but also a detailed description of the coding criteria and its application is provided (see section 3.10.3). It is hoped that by making the data extraction and manual analysis explicit and transparent, further replications of the study can be made.

### 1.7. Structure of the thesis

This thesis is divided into six chapters. Following this introduction, the second chapter provides a historical account of the development of MWVs as well as a description of their form, syntactic and semantic properties, which differentiate them from other types of MWUs. The chapter then goes on to discuss the contributions of corpus linguistics to the study of MWVs in both L1 and L2 contexts and identifies areas where further learner corpus-based research on MWVs is needed. In the remainder of the second chapter, an overview of MWVs and similar verbal structures in the three L1 backgrounds in which MWVs are analysed in the present study (i.e. Chinese, Italian and Spanish) is provided. This overview is followed by a summary of research findings

on the link between task type and MWV use. Chapter 3 is concerned with the methodological aspects of the study. It first presents a description of the selected corpus, the Trinity Lancaster Corpus, and explains the rationale for its selection. Details regarding the process of data extraction and statistical analyses are also specified. Special emphasis is given to the manual coding of the data and the application of the coding scheme designed for this purpose. Chapter 4 presents the results of the study. These are organised according to the research questions that the study aimed to answer. In chapter 5, the findings are analysed in light of three main areas: i) MWV use in the Trinity Lancaster Corpus (i.e. the frequency and coverage of MWVs, lexical verb and particle productivity, non-canonical MWV forms, and polysemy), ii) the role of learner variables (i.e. L2 proficiency and L1 background), and iii) the effect of situational variables (i.e. task type) on MWV production by L2 speakers. In chapter 6 the main findings of the study are summarised. The contributions to theory and methodology are outlined and followed by an overview of the pedagogical implications of the findings. Finally, limitations are discussed and recommendations for further research on MWVs in L2 communication are offered.

# Chapter 2: Literature review

## 2.1. Introduction

The following chapter first gives an overview of the historical development of MWVs and the linguistic changes from which the PV and PPV as we know them today originated. It then goes on to describe the internal structure of PVs and PPVs. Earlier definitions of the term 'phrasal verb' are also contrasted with those proposed by phraseological studies, and arguments are provided to distinguish MWVs from similar multi-word constructions, namely, collocations and idioms. The chapter also provides a description of the syntactic and semantic properties of MWVs as well as the contributions of L1 and L2 corpus linguistic research to the study of MWVs. The last sections of the chapter focus on MWVs as observed in different L1 verbal systems, particularly Chinese, Italian and Spanish, and the role of task type in L2 production of MWVs.

## 2.2. The MWV in English

### 2.2.1. The historical development of the PV and PPV

The PV as it is known today is a structure of common Germanic origin rather than a particularly English one. This means that PVs can be found in a group of genetically related languages including German, Dutch, Swedish, Danish, Icelandic, Norwegian, Yiddish, Faroese, and West Frisian (see section 2.4.1). Interestingly, PVs have also been found in other non-Germanic languages, for instance, Old French (e.g. Dufresne, Dupuis & Tremblay, 2003) and Italian (e.g. Benincà & Poletto, 2006; Iacobini &

Masini, 2006). According to Masini (2005), Italian MWVs emerged probably as a result of either an autonomous internal development in Italian or a process whereby PVs were loaned from previously existing dialect varieties and then introduced into standard Italian. The historical development of PVs in Germanic and non-Germanic languages like Old French and Italian, however, followed different paths, which partly explains the fact that although there are PVs in those languages, not all of them are identical to the PV in English (Thim, 2012).

The modern English PV is the result of a series of morphological and syntactic changes, most notably the placement of the particle after the verb (e.g. *come out*). In Old English, particles were characterised as preverbs, that is, "morphemes that appear in front of a verb, and which form a close semantic unit with that verb" (Booij & van Kemenade, 2003, p. 1). Not only did particles precede verbs then, but they were also attached to them as preverbal prefixes (e.g. *forbærnan, to burn up*), forming what was known as complex verbs (Booij & van Kemenade, 2003). The reasons for the separation of particles from verbs have been reported to be mainly phonological. For example, Gries (2003) explains that both elements (verb and particle) received stress which led to a contradictory situation where there were two consecutive stressed items. The separation was a way to solve the stress placement problem. Although this phonological explanation may account for the detachment of particles, it does not clarify why particles later moved to a post-verbal position. In this respect, Thim (2012) argues that particles were postposed as a result of the positional changes of the different elements in the verb phrase that took place during the rise of the Modern English word order. Such rise occurred when the once dominant object-verb (OV) order shifted to a verb-object (VO) pattern in the Modern English period. According to Bacchielli (1999),

Hiltunen (1994) and Gries (2003), the changes that the PV suffered occurred simultaneously with the synthetic to analytic shift where grammatical and morphological meanings started to be carried by various language items rather than being packed in a single unit (Danchev, 1992). This also implied that the semantics of PVs were carried by two units, a phenomenon that Bolinger (1971) termed "semantic spreading" (p.45) and that Brinton and Akimoto (1999) also found present in complex verbs (e.g. *have a drink, make a call*) and complex prepositions (e.g. *in relation to, by means of*).

In contrast to PVs, historical information regarding the roots of PPVs is rather scarce. This is probably due to the fact that PPVs are considered an offspring of PVs and prepositional verbs and tend to be subsumed under the PV category (e.g. Rodríguez Puente, 2019) or deliberately excluded from historical analyses of MWVs in English (e.g. Hiltunen, 1983). Despite the lack of historical information about PPVs, their development has been reported to be similar to the one of PVs in the sense that their origins also date back to the Old and Middle English periods (Denison, 1985). According to Claridge (2000), the first mention of PPVs as a type of MWV in their own right can be attributed to Mitchell (1958), who proposed one of the first MWV classification schemes known to the present day. The establishment of the term PPV, then, came rather late given that the first attested instances of PPVs date back to Old English. Regarding the reasons that motivated the combination of a verb + particle with a preposition to form PPVs, Denison (1998) argues that PPVs derived from already established PV forms by means of a systematic supplementation process aided by an overall increase in the use of prepositions during the Middle English and Late Modern English periods.

This section provided a brief account of the historical development of PVs and PPVs from Old English to Modern English. On the one hand, like combinations of verb + particle in many other Germanic languages, the English PV transformed from a structure where the verb and its particle were fused together as one unit to a combination where the verb was separated from its particle but still preceded by it. Finally, it evolved into a construction where both elements were two independent orthographic words, with particles taking on a post-verbal position. On the other hand, the English PPV derived from already established PV forms by taking in one preposition. Although the term PPV was not coined until 1958, instances of PPVs in Old and Middle English had already been documented.

### 2.2.2. Defining the term 'phrasal verb'

There is common agreement regarding the surface form of a PV, which consists of the combination of a verb and an adverbial particle[2]. Cappelle (2007), however, argues that the 'verb' in a PV does not necessarily have to be an actual verb, but it can be an element of a different word class with a verbal function. For example, Thim (2012) explains that the word '*terrace*' in the PV '*terrace out'* is not a lexical verb, given that this word never functions as a verb outside of the PV structure. Therefore, it would be more accurate to consider PVs as combinations of any 'verbal element' and a particle.

The verbs that typically participate in the formation of PVs tend to be high-frequency and monosyllabic (e.g. *come, turn*). This certainly does not mean that polysyllabic verb forms cannot appear in PVs, for example, *recipe out, party out*, *Christmas out*, *gyrate*

[2] Adverbial particles are also referred to as adverbs or prepositions in the literature. See Thim (2012) for a discussion.

*over*, and *cherish up* (Weirszycka, 2013; Gardner & Davies, 2018). The particles in PVs typically come from a small group that includes the following items: *aback, about, aboard, above, across, after, ahead, along, apart, around, ashore, aside, astray, asunder, away, back, behind, by, down, forth, forward(s), home, in, off, on, out, over, past, round, through, to, under*, and *up* (Cowie & Mackin, 1975; Quirk et al., 1985; Claridge, 2000; Huddleston & Pullum, 2002). Rather than including new members in this group, research studies tend to exclude some of the previous particles due to their overall infrequency in PV constructions (e.g. *aback*), restrictions in terms of genre (e.g. *asunder* in literary fiction) or their lexical status. Gilquin (2015b), for example, discarded *home* as a particle, arguing that this word has a lexical rather than a grammatical status.

Even when the structure of PVs is relatively clear, their study has been marked by a lack of consensus regarding their linguistic status as single units or phrases. Traditional lexical studies have put forward strong arguments in favour of PVs as being single lexical units. One of such arguments relates to the ability of PVs to participate in morphological derivation processes, a typical characteristic of words rather than of syntactic phrases (Farrell, 2005; Chapman, 2008; Cappelle, 2010). PV derivations can take various forms, the -er nominalisations being the most common ones (e.g. *fixer-upper* from *fix up*, *dropper-in* from *drop in*, *come-outer* from *come out*). Sometimes the -er suffix joins a preposed particle followed by the verb (e.g. *onlooker, bystander*), thus resembling the word order of PVs in Old English (see section 2.2.1). Another type of nominal form is the one in which the verb and the particle are written as one word (e.g. a *dropout* from *drop out*, or a *hangup* from *hang up*). These nominalisations have then been taken as evidence of PVs' word-like behaviour.

Although the morphological argument described above might speak in favour of the status of PVs as single lexical units, PVs should not be regarded as single words. If a word is taken as an uninterrupted sequence of letters, PVs are clearly not single words since they comprise two separated elements (or sequences of letters): a verb and a particle (Thim, 2012). Moreover, if the verb and its particle formed a single word, intervening elements such as pronominal objects or adverbs could not be inserted in between the verb and its particle (Di Sciullo & Williams, 1987; Cappelle et al., 2010). In fact, the presence of intervening elements makes PVs violate the principle of lexical integrity that characterises single words. This principle posits that "the syntax neither manipulates nor has access to the internal form of words" (Anderson, 1992, p.84). What this means is that, as opposed to words, the components of a PV can be operated by syntactic rules and can indeed be separated; therefore, PVs are "certainly lexical units, but not words" (Booij, 2007, p.22). It is also worth noticing that in nominalisations like *fixer-upper* the derivational suffix attaches to both verb and particle making them two distinct elements rather than one lexical unit, which clearly contradicts the argument that PVs have a word-like status.

However, the idea that PVs are single units has found further support in a more recent body of neurolinguistic literature. Neurophysiological studies have suggested that PVs are processed and accessed as single lexemes, which evidences their word-like status (Kapatsinski & Radicke, 2009). Cappelle et al. (2010), for instance, used brain imaging to analyse native speakers' responses to particles in existing PVs (e.g. *cool down*) and in infelicitous combinations (e.g. *cool up*) in order to determine whether PVs were syntactically assembled or acquired as single lexical units. Their investigation was based on the neurophysiological fact that single words show enhanced brain responses

as opposed to word strings which are syntactically linked and show reduced brain responses. The results of this study revealed that brain responses were enhanced in the presence of particles in existing PVs, thus providing evidence that "in the language of the brain" (p. 197) the verb and particle in a PV are processed and stored as a whole lexical chunk. Moreover, Cappelle et al. (2010) found no correlation between the lexical status of PVs and their degree of transparency. In other words, PVs were accessed as holistic lexical units regardless of their semantic type, i.e. opaque or transparent. Neurolinguistic studies have certainly made an important contribution to the description of PVs by showing that they are stored in the mind as unitary wholes in almost the same way that single verbs are stored (Matlock & Heredia, 2002). However, the fact that PVs resemble words in terms of processing and mental storage does not necessarily make them single-word units just as many other types of MWUs (e.g. collocations) that have also been found to be stored and retrieved as wholes are not single words (Siyanova-Chanturia & Van Lancker Sidtis, 2019). What PV storage and processing as single verbs rather shows is that there exists a strong link between the verb and its particle.

PVs have also been regarded as phraseological units, rather than as words only, in the tradition of corpus-based and phraseological research. The arguments behind this position relate to the fact that PVs display syntactic and semantic characteristics that are typical of phrases rather than of words, for example, polylexicality, fixedness, and (non-) compositionality. Because of these properties, any attempt to define PVs simply as words would not do justice to them or, as Sinclair (2004) argues, to any phraseological unit.

In an effort to define verb-particle combinations while taking their syntactic and semantic properties into account, phraseological corpus-based studies have described PVs as one type of MWV. MWVs include phrasal, prepositional and phrasal prepositional verbs[3]. Table 2.1 contains a set of sentences illustrating the internal structure of these verb types. Other polylexical verbal forms that are sometimes classified as MWVs are sequences of a verb + noun phrase (+ preposition) (e.g. *take a look at*), verb + prepositional phrase (e.g. *take into account*), and verb + verb (e.g. *make do*) (Biber et al., 1999).

Table 2.1 Structure of MWVs

| **MWV type** | **Internal structure** | **Examples** |
|---|---|---|
| **PV** | verb + particle | ***Come on**, tell me about Nick!* (Biber et al., 1999) |
| **Prepositional verb** | verb + preposition | *The first goal **came from** Tim Cliss.* (Biber et al., 1999) |
| **PPV** | verb + particle + preposition | *There was no telling what a girl like that might **come up with**; they might beat Mavis Bramley yet.* (Claridge, 2000) |

[3] Other classifications of MWVs have been proposed. See e.g. Fraser, 1976; Denison, 1981; and Claridge, 2000.

Given that PVs and prepositional verbs are often superficially alike (e.g. *run up a bill* vs. *run up a hill*), a series of syntactic, semantic and phonological tests have been proposed to distinguish both verbal forms from each other. These tests are summarised in Table 2.2. Albeit informative, none of the tests is considered infallible or free of exceptions (O'Dowd, 1998), which makes their application problematic. However, Biber et al. (1999) concede that two tests, i.e. subject-verb inversion and wh- question formation, are relatively reliable and sufficient to distinguish between most PVs and prepositional verbs.

Table 2.2 Tests to distinguish PVs from prepositional verbs

| *Type* | *Test* | *Examples* | *Reference* |
|---|---|---|---|
| **Only PVs allow:** | | | |
| **Syntactic** | Passivization | Aunt Ada brought up Roy. → Roy was brought up by Aunt Ada.<br>We turned off the road. → *The road was turned off. | Quirk et al. (1985) |
| | Particle movement | We turned off the light. → We turned the light off.<br>We turned off the road. → *We turned the road off. | O'Dowd (1998) |
| | Personal pronoun insertion | They called him up. → *They called up him.<br>They called on him. → *They called him on. | Quirk et al. (1985) |
| | Verb nominalisation | He looked up the information. → His looking up of the information.<br>He looked at the window. → *His looking at of the window. | Bolinger (1971) |
| **Semantic** | One-word verb substitution | They have to call up that man. → summon<br>I left those out. → omitted | Quirk et al. (1985) |

| | | | |
|---|---|---|---|
| **Phonological** | Stress placement on the particle | She switched ON the light. → The light was switched ON.<br>He CALLED on the dean. → The dean was CALLED on. | Quirk et al. (1985) |
| **Only prepositional verbs allow:** | | | |
| **Syntactic** | Adverb insertion | The pilot jerked the lever right back.<br>*They put the meeting hurriedly off. | Quirk et al. (1985) |
| | Subject-verb inversion | Back moved the man.<br>*Up blew the tank. | Quirk et al. (1985) |
| | Preposition replacement | She took in the box. → She took aside/outside/down the box.<br>She took in her parents. → *She took outside/inside her parents. | Quirk et al. (1985) |
| | Fronted particles in question formation | He called on the dean. → On whom did he call?<br>He turned the suspect in. → *In whom did he turn? | Quirk et al. (1985) |
| | Syntactic versatility | On whom did he call? → On his mother | Quirk et al. (1985) |

| | | |
|---|---|---|
| | Whom did he call on? → On his mother<br>Did he call on the dean OR on his friend?<br>He calls on the dean MORE OFTEN THAN on his friend. | |
| Placement of prepositions before relative clauses | The man on whom they called.<br>*The man up whom they called. | Quirk et al. (1985) |
| Verb-gapping | He sped up the street, and she, up the alleyway.<br>*He sped up the process, and she, up the distribution. | O'Dowd (1998) |
| Noun phrase ellipsis | We turned off (the road).<br>*We turned off (the light). | O'Dowd (1998) |
| Conjunction-reduction | We turned off the road and onto the highway.<br>*We turned off the light and on the stereo. | O'Dowd (1998) |

Defining PVs from a phraseological perspective, however, is not without its problems. First, although PVs are considered MWVs, they are often excluded from empirical research on MWV forms. For example, Moon (1998) excluded PVs in her study on phrasal lexemes "for practical reasons" (p.79) without specifying the rationale behind that decision. In their 2004 study, Grant and Bauer decided not to focus on PVs on the basis of the fact that "they are such a large group of MWUs [multi-word units] that they merit separate and thorough research of their own" (p.39). Second, since there is a wide variety of MWUs whose boundaries still remain fuzzy (Cowie, 1998), phraseological and corpus-based studies disagree as to which specific type of MWU PVs are. As a result, it is common to see that PVs are treated as grammatical collocations or idioms, two different types of MWUs. Granger & Paquot (2008), for example, include PVs in the referential phraseme category of their MWU classification together with both idioms and grammatical collocations. It is beyond the scope of this section to fully describe collocations and idioms. However, the main characteristics of these two units will be used to contrast them with PVs.

Grammatical collocations or colligations are defined as the mutual company (co-occurrence) of a lexical word and a grammatical word (e.g. verb/noun/adjective + preposition) (McEnery, Xiao & Tono, 2006; Granger & Paquot, 2008; McEnery & Hardie, 2012). The assumption behind this is that the verb in a PV is the lexical word and the particle is its grammatical company. Lipka (1972) and Sroka (1972) call PVs verb-particle collocations, yet they do so for different reasons. Lipka (1972) prefers the term 'collocation' because it does not presuppose dominance of the verb or particle in the relationship, i.e. the verb is not responsible for selecting its particle and vice versa. In contrast, Sroka (1972) argues that the term 'collocation' does not "involve

restrictions concerning the 'unity' of the verb and particle" (p. 22). By restrictions, he means that adverbial particles tend to have a stronger connection with the verb in comparison with prepositions. However, these restrictions do actually exist and have been well-documented (Quirk et al., 1985; Claridge, 2000; Dehé, 2002). In his book The English Verb, Palmer (1988) states that PVs are indeed subject to several restrictions. One of them relates to the limited set of particles that can join verbs in PV constructions. More recently, Waibel (2008) has justified the use of 'grammatical collocation' to refer to PVs by arguing that the term "implies that phrasal verbs are at the interface of grammar and lexis since they consist of one open-class item (the verb) and one closed-class item (the particle)" (p.18). Interestingly, Waibel (2008) calls PVs 'collocations' and 'idioms' interchangeably in the same section of her book on German and Italian speakers' use of PVs.

An idiom consists of a conventionalised expression which is often non-compositional, i.e. its meaning generally cannot be deduced from the individual meanings of its components (Fernando, 1996; Wulff, 2008). Since PVs often show "a similar range of idiomaticity types to fixed expressions and idioms" (Moon, 1998, p.3), they have been called lexemic idioms. They are characterised by the fixed order of their components, an aspect that they share with most idiomatic expressions (Makkai, 1972; Alexander, 1987). Cowie (1993) points out that there are idiomatic PVs and non-idiomatic verb-particle combinations that cannot be considered PVs. The first ones correspond to semantically opaque PVs whereas the second ones have ordinary, transparent meanings. In other words, idioms include non-compositional PVs but exclude all other semantic types of PVs.

The debate arising from whether PVs can be classified as collocations or idioms seems to relate to the way those two terms have been used as a function of the semantic compositionality of the PV. It has been suggested that transparent PVs, where a verb and its particle retain their individual meanings, are collocations whereas non-transparent PVs are idioms (Cowie, 1993). Waibel (2008) also pursues this line of argument by claiming that "even if the criterion of idiomaticity does not apply, phrasal verbs can be considered phraseological units, by analogy with (transparent) collocations" (p. 18). The lack of compositionality, therefore, seems to be the primary criterion to determine whether a PV is a collocation or an idiom. This raises the question of how semi-compositional, i.e. aspectual, PVs would be classified. In addition, if transparent PVs are indeed grammatical collocations, it is unclear whether the verb is the node and the particle its collocate. Even if this were so, it is difficult to determine the extent to which a verb actually collocates with a particle since particles do not exist outside the PV. Words like *out, up, off,* or *in* become particles only when being part of a PV. Outside PVs, particles stop behaving 'particle-like' and adopt their prepositional function. In the present study, PVs and PPVs are considered phraseological units and, as Thim (2012) says, "not 'just' 'words'" (p.55). Neither are PVs or PPVs treated as idioms or collocations. Although these are, arguably, relatively general labels, the terms 'phraseological unit' and 'multi-word unit' are preferred as they acknowledge the syntactic and semantic characteristics of PVs and PPVs.

### 2.2.3. Defining the term 'phrasal-prepositional verb'

As shown in Table 2.1, PPVs consist of three-word combinations formed by a verb + particle + preposition[4]. In other words, they consist of PVs with a preposition added to them. The lexical verbs and particles functioning in PPVs tend to be same as those in PVs. Even though there seem to be no restrictions as to the types of prepositions that can join verbs and particles to form PPVs, the ones typically occurring in those combinations tend to be part of a rather small set: *from, to, with, on, of, for, at*, and *in* (Quirk et al., 1985; Biber et al., 1999). While the form of PPVs is relatively clear, little has been said about the internal semantic unity that characterises them. One notable exception is Claridge (2000), who adopts a semantic approach to diachronically analyse PPVs and argues that their three components (i.e. verb + particle + preposition) should display internal semantic cohesion so that the combination can actually be considered a PPV (e.g. *put up with somebody)* and not just a PV randomly followed by a prepositional phrase (e.g. *walk up with somebody)*.

The internal semantic unity of PPVs, however, has been described to depend on the existence of one-word equivalents. Put differently, a verb + particle + preposition sequence is considered a PPV form provided that it can be paraphrased as one word (Claridge, 2000). As a case in point, *put up with* can be paraphrased as *tolerate* whereas *walk up with* seems to lack a one-word equivalent and thus qualifies as a PV plus a free preposition. Establishing the semantic unity of PPVs on the basis of one-word paraphrases is problematic though. Many instances of PPVs and PVs lack synonymous one-word counterparts (Celce-Murcia & Larsen-Freeman, 1999; Fletcher, 2005; Thim,

[4] Scholars like Quirk et al., (1985) and Palmer (1988) also call the preposition in PPVs a particle.

2012). Moreover, even in cases where PPVs seem to have one equivalent, the PPV and its apparent one-word paraphrase can only be regarded as equivalents in very general terms, with the one-word paraphrase failing to express the full meaning of the PPV form. For example, Biber et al. (1999) provide *anticipate* as a synonym of *look forward to*. Although *anticipate* is indeed related to the meaning of *look forward to*, it does not convey the idea of 'awaiting eagerly' that is implied in the PPV.

Perhaps a stronger sign of semantic unity between the components of a PPV form is the one mentioned by Denison (1981). He argues that the unitary semantic status of PPVs is related to the fact that their components are unlikely to be interrupted or inverted. However, it should be pointed out that a few PPVs are complex transitive, which means that they can take an intervening element in between the particle and the preposition (e.g. *take out your anger on me*) and between the verb and the particle (e.g. *take it out on me*). Despite these relatively few cases, the presence of intervening elements and inversion are highly unusual (Palmer, 1988; Biber et al., 1999; Claridge, 2000), which supports the argument for the semantic unitary status of PPVs.

### 2.2.4. Syntactic and semantic properties of MWVs

The criteria for the identification of MWVs have been well documented. This section first reports on a common core set of syntactic and semantic characteristics of PVs and PPVs which have been used to identify them. Second, it comments on the different positions regarding transparency in PVs. Table 2.3 below provides an overview of the syntactic and semantic characteristics of MWVs. Four pieces of information are specified in the table: i) the type of syntactic or semantic property, ii) an explanation of how the property applies to PVs and PPVs, iii) sample sentences in which PVs and

PPVs illustrate those properties, and iv) reference works from which the sample sentences were taken and in which the specific property has been discussed.

Table 2.3 Syntactic and semantic properties of PVs and PPVs

| | **Type** | **Explanation** | **Example(s)** | **Reference(s)** |
|---|---|---|---|---|
| **Syntactic properties** | Transitivity | -PVs can be intransitive or transitive. If transitive, PVs can take nominal or pronominal direct objects. Short noun phrases can come before or after the particle. | -*The whole house* ***blew up***.<br>-*My father has to* ***turn off*** *the alarm*.<br>-*My father has to* ***turn*** *the alarm* ***off***. | Bolinger (1971), Quirk et al. (1985), Jackendoff (2002) |
| | | -A few PVs can occur in ditransitive forms. | -*Valerie* ***packed*** *her daughter* ***up*** *a lunch.* | Dehé (2002) |
| | | -PPVs are mainly monotransitive. The complement of the preposition in a PPV acts as its direct object. | -*Oh I shall* ***look forward to*** *this now*. | Biber et al. (1999) |
| | | -A few complex transitive PPVs can take two objects. | -*We* ***put*** *our success* ***down to*** *hard work*. | Quirk et al. (1985) |

| | | | | |
|---|---|---|---|---|
| | Separability | -The separation of the verb and particle in PVs is mandatory when the direct object takes the form of a reflexive or object pronoun.<br><br>-Lengthy noun phrases acting as direct objects preferably occupy the post-particle position.<br><br>-In PPVs, verbs and particles are not separated unless the PPV is one that can take two objects. | -*Can you **put** it **out**?*<br>-*Can you **put out** it?**<br><br>-*Lila **looked up** the answer to the question that was on everyone's mind.*<br><br>-*Oh I shall **look forward to** it.*<br>-*He **fobbed** me **off with** a feeble excuse.* | Quirk et al. (1985), Celce-Murcia & Larsen-Freeman (1999)<br><br>Quirk et al. (1985)<br><br>Biber et al. (1999)<br>Palmer (1988) |
| | Transparency | -The meanings of PVs range on a cline of compositionality. Three semantic categories are identified:<br>*a) Compositional*<br>Both verbs and particles are transparent in meaning. Particles indicate direction or location. | -*Please, **come in**.* | Quirk et al. (1985), Huddleston & Pullum (2002) |

| | | | | |
|---|---|---|---|---|
| **Semantic properties** | | *b) Semi-compositional*<br>The verb keeps its meaning but not the particle. Particles function as aspectualisers or telic aktionsart markers.<br><br>*c) Non-compositional*<br>The meaning of the combination cannot be deduced from any of the individual meanings of its components.<br><br>-The meanings of PPVs also range on a cline of compositionality; however, fully compositional (i.e. transparent) PPVs tend to be rare. | *-She **used up** all the film.*<br><br>*-They **turned in** the suspect to the police.*<br><br>*-The doctor's voice **broke in on** her thoughts.*<br>*-If I stretch my imagination, I can admit to feeling a little tired lately, but **put** that **down to** the ageing process.* | Claridge (2000) |
| | Polysemy | -One PV or PPV form can have multiple meaning senses associated with it. | *-I couldn't **take** it all **in** at once.*<br>*-The army **took in** the rest of the city.*<br>*-Teenagers were **taken in** and later recruited to work in the mine.* | Biber et al. (1999) |

| | | | | |
|---|---|---|---|---|
| | | -The various meaning senses are not necessarily completely different from one another but may be related. | -*You can **put** the lunch **down to** my account*.<br>*-I did notice that you were depressed, but I **put** it **down to** drink*. | Cowie & Mackin (1993) |

### 2.2.5. PVs and free combinations

There has been a discussion about a particular semantic type of PVs, the literal verb-particle combination, whose meaning is fully transparent and thus might resemble verb + preposition sequences. Verbal forms like *'go out', 'go back', 'come out',* or *'come back'* have often been referred to in the literature on MWVs as 'free combinations', a term that can probably be attributed to Quirk et al. (1985) who define 'free combinations' as structures in which "the verb and the adverb have distinct meanings on the other" (p.1152). Biber et al. (1999) and Dixon (1982) also draw this distinction between free combinations and PVs on the basis of three criteria. First, it is argued that substitutions of verbs and particles are possible in literal combinations as opposed to semi- and non-compositional PVs. Second, PVs are syntactically cohesive and do not welcome the insertion of adverbs whereas free combinations do. Third, word order alterations are allowed in free combinations and not in PVs.

However, the classification of literal PVs as free combinations in linguistic studies poses, to my view, two problems. First, it is done on fuzzy semantic grounds. Authors repeatedly acknowledge the difficulty of drawing clear-cut boundaries between the three semantic types of PVs (i.e. compositional, semi-compositional and non-compositional) (see Table 2.3). Second, the classification disregards the fact that literal PVs behave syntactically like semi- and non-compositional PVs (Hampe 2002; Rodríguez-Puente, 2019). Thim (2012) states that literal PVs are the only combinations that exhibit the "full range of syntactic properties typical of phrasal verbs" (p. 15) and emphasises the lack of clear explanations from advocates of the PV vs. free combination distinction as to why the syntactic properties of PVs only characterise those that are non-compositional and not literal combinations. Moreover, as observed by Rodríguez-

Puente (2012), literal combinations are the most basic type of PVs and actually in many cases they are "the source for the formation of figurative and non-compositional combinations" (p.73). Based on these arguments, the distinction between PVs and free combinations appears rather problematic and difficult to justify. The studies that follow this distinction may, as a result, omit an important set of PVs.

## 2.3. The role of corpus linguistics in MWV studies

As part of his recommendations for further research on MWVs, Bolinger (1971) calls for a need to conduct lexical surveys, given that "no large-scale count of phrasal particles and verbs entering into phrasal combinations has been made" (p. 175). Since the early 1970s, corpora (i.e. large collections of naturally-occurring language) have allowed researchers to conduct this type of large-scale analysis of MWVs, with findings from corpus-based studies having applications that range from MWV description and pedagogy to dictionary making and materials design. The purpose of this section is to review the main contributions of L1 and L2 speaker corpora to the study of MWVs and to highlight the suitability of corpus linguistics as a methodology to examine MWVs.

### 2.3.1. MWVs in L1 speaker corpora

#### *2.3.1.1. Frequency*

Perhaps not surprisingly, the most important contribution of corpus linguistics to the study of MWVs is providing information about their frequency, yet the concept of frequency has been operationalised in different ways, including frequency of verbs in MWV constructions, particles, and meanings. Overall, L1 corpora have shown that

MWVs are highly frequent in English (Sinclair, 1991; Gardner & Davies, 2007; Diemer, 2014), with Biber et al. (1999) reporting that on average PVs occur 2,000 times and PPVs 400 times per million words, particularly in spoken registers. However, MWVs have not always been that common as diachronic analyses of historical corpus data have revealed. In a multi-corpus study, Rodríguez-Puente (2016) shows that there is a tendency for MWVs to increase over time and that MWVs grew in number particularly from the 1700 onwards. Nevertheless, rather than showing a steady growth, the development of MWVs was non-continuous and actually slowed down in the 1800-1850 period (Brown & Palmer, 2015; Rodríguez-Puente, 2016).

Rather than discovering syntactic or semantic properties of MWVs that were not known from previous studies, L1 corpora have provided solid evidence to support claims about MWV behaviour in English which were previously based on researchers' intuition. For example, earlier studies on MWVs (e.g. Bolinger, 1971 & Sroka, 1972) speculated that certain particles are more prolific than others. The development of more sophisticated corpus search tools has allowed not only frequency counts of the verbs and particles that typically participate in the formation of MWVs but also the identification of the most productive verbs (e.g. *go*, *come*) and particles (e.g. *out*, *up*) in L1 English (Liu, 2011). Moreover, L1 corpus-based studies have found that although MWVs are highly frequent, they are not necessarily formed by a wide variety of verbs and particles. This was demonstrated, for example, by Gardner and Davies (2007), Liu (2011), and Lee (2015), who found that a small group of verb lemmas combined with an also small number of particles accounts for the majority of MWV occurrences in the 1994's British National Corpus (BNC) (British National Corpus Consortium, 2000), the Corpus of

Contemporary American English (COCA) (Davies, 2008), and the Michigan Corpus of Academic Spoken English (MICASE) (Simpson et al., 2002).

*2.3.1.2. Polysemy*

In addition to providing information about frequency of form, corpus linguistic research has looked at the frequency of meanings associated with a single polysemous MWV form. It is well-known that MWVs are highly polysemous (Side, 1990; Darwin & Gray, 1999; Consigny, 2001; Wild, 2011). In fact, Gardner & Davies (2007) report that the most frequent MWVs have 5.6 different meaning senses on average. Not only has corpus linguistics attested the polysemous behaviour of MWVs, but it has also recently shown that the meaning senses of a single polysemous MWV are not equally frequent (Garnier & Schmitt, 2015; Liu & Myers, 2018), a finding that has important pedagogical implications. What is more, corpora have helped to track down the historical development of MWV and particle meanings, which has shed light on the linguistic processes that were responsible for the emergence of those meaning senses. For example, Rodríguez-Puente's (2012) examination of three historical corpora evidences the tendency for non-compositional MWV meanings to derive from literal meanings by means of one of the following processes: addition of aspectual particle + metaphorization of the phrasal compound, multiple metaphorization, specialization, reduction of elements, creation of brand-new combinations, and changes in the real world. The last process refers to how the meanings of some MWVs change through time in order to adapt to new technological developments (e.g. *log in, back up*).

Corpus-based studies that have focused on specific MWV forms have further contributed to the study of particle meaning and its development (e.g. Glodovic, 2014;

Mahpeykar & Tyler, 2015). Adopting a historical corpus-based cognitive analysis, Ishizaki (2012) draws on data from the Corpus of Early English Correspondence Sampler (CEECS) (Nurmi, 1999) and the Penn Parsed Corpus of Modern British English (PPCMBE) (Kroch, Santorini & Diertani, 2010) in order to examine the aspectual and idiomatic properties of the particles *away* and *out* in Early and Late Modern English. Having access to a large data sample allows Ishizaki to observe that MWVs containing *away* and *out* have gone through an idiomatisation process in which the particles seemed to have lost their concrete and literal meanings. In a similar vein, Wild (2010) and more recently Leone (2016) provide detailed historical accounts of the semantic features of MWVs containing the particles *back*, *down*, *out*, *up*, and *on*. In both corpus-based investigations, the authors looked at the linguistic processes that shaped the semantic evolution of those particles and the MWVs in which they appear. They also provide details as to the approximate time periods when the meanings of the MWVs containing those particles shifted from spatial to the more aspectual and idiomatic meanings that characterise their present-day use. As opposed to early diachronic studies, corpus-based research of the kind described above has provided a comprehensive description of both MWVs and the recurrent usage patterns associated with MWVs in different periods.

*2.3.1.3. Register and genre distribution*

Regarding the use of MWVs in discourse, corpus insights have been of particular importance because they have provided information about MWV register distribution. In this respect, the study conducted by Biber et al. (1999) using the Longman Spoken and Written (LSWE) corpus (Biber et al., 1999) is worth noting. As a result of their analysis of four sub-corpora, each representing a different register in British and

American English, Biber and his colleagues demonstrate that MWVs occur more frequently in conversation and fiction than in news and academic prose. They attribute this finding to the informal tone conveyed by MWV constructions which tends to characterise spoken interaction. Their analysis also brought to light the distribution patterns of MWVs across a variety of semantic domains (e.g. activity, mental, communication, occurrence, and aspectual) and transitivity patterns. Two important findings resulted from this analysis. First, the most common MWVs come from the domain of activity MWVs, i.e. verbs expressing events or actions performed by a usually human subject (e.g. *stand up, put on, hold on to*). Second, conversation and fiction have a heavier reliance on intransitive PVs (e.g. *come on, shut up*) as opposed to news and academic prose which had fewer than 40 occurrences of those verbs per million words. Despite the fact that MWVs were not the primary focus of Biber et al.'s work, their findings reveal information about the communicative purposes behind the use of MWVs. For example, as observed by Biber et al. (1999), in conversation speakers often talk about others' actions or their own, hence the prevalence of activity intransitive PVs. Also, speakers tend to use intransitive PVs which commonly denote imperative actions, again a characteristic of spoken discourse.

Biber et al.'s (1999) study of MWVs is not without its limitations though. First, MWVs that did not meet the established criteria (40 times per million words in at least one register) were not included in the study. Second, one can also argue that grouping all kinds of conversation in one register is an oversimplification given that there may be different types of conversation whose levels of formality can vary. Finally, the classification of verbs in semantic domains was based on core meanings, thus it failed to consider extended meaning senses. Even so, the work of Biber and his colleagues

shows the enormous potential of corpus linguistics to describe the usage-based properties of MWVs. It also represents the first attempt to examine the semantic domains of MWVs from a corpus perspective. In an earlier study, cognitive linguist Rudzka-Ostyn (1988) analysed a group of MWVs from the domain of verbal communication; however, her approach did not involve corpus methods.

While the study conducted by Biber and colleagues is illuminating in terms of MWV use in present-day English, other types of L1 corpus-based studies have made it possible to observe the historical distribution of MWVs in a wider range of genres. Perhaps more importantly, the use of corpora in these studies has helped to account for the reasons why MWVs have been traditionally common in spoken genres and have become more frequent in some written genres during different time periods (e.g. Akimoto, 1999; Kytö & Smitterberg, 2006). According to Smitterberg (2008), MWVs have historically occurred more frequently in what he calls non-expository genres (e.g. drama, fiction, letters, and trials), which are to some extent connected to spoken or informal language. However, as his corpus analysis demonstrated, MWVs started to become increasingly more common in written genres (e.g. history) during the nineteenth century. He explains these findings in terms of the colloquialisation of written genres that took place during that period. Put differently, written genres started to adopt linguistic features characteristic of conversational speech, which led to an increase in the use of MWVs. Drawing on data from a different historical corpus, Rodríguez-Puente (2019) went one step further and investigated the distribution of MWVs across i) a wider variety of formal and informal registers (e.g. diaries, news, science, sermons, journals, drama, medicine, etc.), ii) American and British English varieties, and iii) time periods covering from 1650 to 1990. Taken together, the results of her study showed that,

despite genre variation, MWVs have tended to feature more prominently in speech-related genres as early as the eighteenth century.

Despite the well-known limitations of historical corpora (e.g. some spoken genres might not faithfully represent spoken language) (Rodríguez-Puente, 2019), corpus analyses like the ones described above have contributed greatly to the study of register and genre distribution of MWVs. They have made it possible to automatically search a vast amount of historical data from the last four hundred years not only to identify patterns and variation in the distribution of MWVs across spoken and written registers but also to throw light on the linguistic processes that were responsible for such variation.

### 2.3.2. MWVs in L2 speaker corpora[5]

#### *2.3.2.1. Frequency*

Arguably, the most significant contribution of L2 corpus-based analyses to the study of MWVs in L2 contexts is providing robust, quantitative evidence for claims about L2 speakers' use of MWVs. In terms of MWV frequency, L2 corpus evidence has helped to confirm two important claims. First, the L1-L2 contrastive analysis approach (see section 1.3.1) adopted by most L2 corpus linguistic research on MWVs has attested that MWVs generally tend to occur much less frequently in learner language than in L1 speaker language (Mondor, 2008; Waibel, 2008; Akbari, 2009; González, 2010; Gilquin, 2011; Weirszycka, 2013; Gilquin, 2015b). What is more, the differences

[5] This section reports on the major contributions of learner corpus-based rather than corpus-informed analyses (e.g. Siyanova & Schmitt, 2007; Schmitt & Redwood, 2011; Becker, 2014; Gardner & Schmitt, 2016) to the study of MWVs.

between MWV frequency in L1 and L2 speaker production tend to be relatively large, with González (2010) and Gilquin (2015b) reporting that MWVs were twice as frequent in the reference L1 corpora that they looked at (i.e. BNC, the Louvain Corpus of Native English Conversation [LOCNEC], and the Louvain Corpus of Native English Essays [LOCNESS]) (De Cock, 2004; Granger, 1998a) than in the learner corpora that they explored (i.e. the International Corpus of Learner English [ICLE] and the Louvain International Database of Spoken English Interlanguage [LINDSEI]) (Granger et al., 2009; Gilquin, De Cock & Granger, 2010). Second, MWV frequency in L2 production has been found to be affected to some extent by the L1 background of the speakers. In this respect, learner corpora have helped to confirm what previous elicitation research (e.g. Dagut & Laufer, 1985; Liao & Fukuya, 2004; Siyanova & Schmitt, 2007) claimed regarding MWV use and L1-L2 structural distance: the absence of MWVs or parallel verbal forms in a speakers' L1 may result in a lower frequency of MWVs in L2 production. For example, in a study combining learner corpus techniques and cognitive linguistic approaches, Gilquin (2015b) investigated MWV constructions in the French components of the LINDSEI and ICLE corpora. Her findings indicate that French speakers used considerably fewer MWVs compared to L1 speakers from the LOCNEC and LOCNESS corpora, which might be related to the fact that French lacks MWVs or similar structures. Drawing on Spanish and Swedish data from the ICLE corpus, González (2010) also found a correlation between MWV frequency and speakers' L1, which he explained in terms of crosslinguistic differences. The lower MWV occurrence in Spanish than in Swedish production was attributed to the apparent absence of MWVs in Spanish (cf. section 2.4.4).

*2.3.2.2. Idiomaticity*

In addition to the overall frequency of MWV occurrence, learner corpus research has made it possible to observe whether there is any correlation between the degree of idiomaticity of MWVs and their frequency. While previous non-corpus-based investigations argued that L2 speakers underuse non-compositional (i.e. idiomatic) MWVs (e.g. Dagut & Laufer, 1985; Hulstijn & Marchena, 1989), learner corpus studies have revealed a much more complex picture with respect to the degrees of compositionality and uses of MWVs (e.g. Mondor, 2008). Drawing on the Polish component of the LINDSEI corpus, Weirszycka (2013) hypothesised that the frequency of MWVs was inversely proportional to their degree of idiomaticity. In other words, L2 speakers would tend to avoid idiomatic MWVs and rely more on transparent ones. Contrary to what was predicted, corpus findings showed that while Polish speakers did use mostly transparent MWVs, the frequency of idiomatic MWVs was higher than that of semi-transparent (i.e. aspectual) MWVs, leading Weirszycka to conclude that "Polish speakers break the linearity at the level of the semi-transparent category" (p. 89). In a more recent study, Deshors (2016) arrived at similar conclusions regarding aspectual MWVs posing more difficulty to L2 speakers than idiomatic MWVs. Deshors (2016) adopted a co-varying collexeme analysis – a corpus-linguistic method from the collostructional analysis family – to ascertain the extent to which particular verb + particle combinations associate with specific semantic uses in two L2 English variants. Put simply, her study measured whether German and French speakers used strongly associated verb-particle pairs idiomatically, aspectually or literally. Her corpus inquiry showed that aspectual MWVs exhibited the weakest associations, thus suggesting that semi-transparent MWVs might be harder to acquire by L2 speakers.

Both Weirszycka's (2013) and Deshors' (2016) works deserve praise for being two of the few corpus-based, rather than corpus-informed, studies directly using learner corpus data to observe MWV idiomaticity in learner language, a complex task on all fronts given the well-known difficulty to classify MWVs into semantic categories. In Weirszycka's (2013) study, the approach to semantically classify the MWVs into transparent, semi-transparent and idiomatic ones appears somewhat unclear in one main respect though. Particles are regarded as the main element defining the semantic category of a MWV on the grounds that the particle is more semantically charged than the verb. Using particles as semantic markers might be relatively easy with literal and semi-transparent MWVs in which the particle tends to contribute a directional (e.g. *fall down*) or aspectual meaning (e.g. *use up*). Nonetheless, it is unclear how the particles in idiomatic MWVs were treated in Weirszycka's classification as both verb and particle equally contribute an idiomatic, semantically opaque meaning (e.g. *put off* meaning to postpone).

*2.3.2.3. Register and genre distribution*

Compared to other areas that have benefited from learner corpus insights, the contributions of L2 corpus research to the investigation of the distribution of MWVs in spoken and written registers have been more modest yet very valuable. The reason why MWV register distribution has not been widely explored using learner corpora is largely due to the fact that the majority of learner corpus-based studies on MWVs have looked at written L2 production only, with argumentative essays from the ICLE corpus (e.g. Nesselhauf, 2005; Mondor, 2008; Waibel, 2008; González, 2010; Deshors, 2016) and other written learner corpora (e.g. Akbari, 2009; Ke, 2013; Chen, 2013, 2018) being the

main source of data in such studies[6]. To the best of my knowledge, five L2 corpus-based studies have investigated MWVs as used in L2 spoken communication (i.e. De Cock, 2005; Gilquin, 2011, 2015b; Märzinger, 2013, and Weirszycka, 2013). Of those five studies, Gilquin's (2011) and (2015b) investigations have directly compared the distribution of MWVs in spoken and written registers. More specifically, her studies examined MWVs in argumentative essays from the ICLE corpus and informal conversations from LINDSEI. Similar findings are reported in both studies in the sense that MWVs were found to occur less frequently in L2 speech than L2 writing, which differs considerably from the register distribution patterns found in L1 speaker communication (Biber et al., 1999). The lower MWV frequency in L2 speech was linked to two factors: L2 speakers' unawareness of the informality conveyed by most MWVs and the less automatic production of MWVs when planning is not possible.

Regarding MWV frequency in interactive spoken registers, Märzinger (2013) examined the distribution of ten highly frequent MWVs across three spoken domains (i.e. educational, leisure and professional) and a variety of speech-related events (i.e. conversation, interview, meeting, panel, press conference, question-answer session, seminar discussion, service encounter, working group discussion, and workshop discussion). MWVs were found to occur more frequently in seminar discussions and in the professional domain, which Märzinger attributed to the informality of the language usually exhibited in both. Märzinger's study is informative and touches upon an area that has been little explored in previous L2 corpus research (i.e. spoken register distribution of MWVs); however, it suffers from three fundamental shortcomings. First,

[6] For a comprehensive cross-varietal investigation of the written register variation of MWVs in world Englishes, see Zipp & Bernaisch (2012). While their study does not rely on learner English data, it offers valuable information about MWVs in different English varieties (e.g. Indian and New Zealand English).

the study was based on a then POS-untagged version of the Vienna-Oxford International Corpus of English (VOICE) (Seidlhofer et al., 2013). As acknowledged by the author throughout the study, the lack of a tagged corpus made it impossible to search for all MWVs which has implications for the validity of the results obtained. Second, the top ten MWVs that were used as the basis of the search came from Gardner & Davies' (2007) list of top MWVs in the BNC in the hope that these same verbs were highly frequent in the L2 corpus. It should be pointed out, however, that Gardner & Davies' (2007) list does not distinguish between top MWVs in spoken and written production and that written registers make up the majority of the BNC data. As a result, the top ten MWVs in the BNC may not have been equally frequent in VOICE. Third, while Märzinger notes that MWVs featured more prominently in seminar discussions, she does not provide a full account of the reasons why MWVs were not as frequent in other speech events that seemed to be just as or even more informal in nature (e.g. working group discussion, workshop discussion, and meeting).

*2.3.2.4. Non-standard MWV forms*

Non-standard MWVs in L2 production have been treated either as errors or lexical innovations, depending on factors such as whether or not the MWV form produced by an L2 speaker or L2 group systematically occurs in other L1 or L2 English varieties or is conventionalised to some degree (Van Rooy, 2011; Deshors, Götz & Laporte, 2016). Not only have L2 corpora allowed the identification of non-standard MWVs in L2 production but they have also provided a description of the actual use of these MWVs in context. Corpus-based qualitative analyses of non-canonical MWVs have proven particularly revealing by showing a variety of patterns of non-standard MWV use,

which were not known from previous elicitation studies tightly controlling for L2 speaker output.

The most often reported patterns of non-canonical MWV use are particle misplacement and the presence of a redundant particle[7] (De Cock, 2005; Nesselhauf, 2005; Gilquin, 2015a). Particle misplacement involves producing a verb + particle combination where a simple verb is required (e.g. *walk up* instead of *walk* or *find out* instead of *find*). Gilquin (2011) explains that redundant particles are those used together with verbs that already express the directional or aspectual meaning of the particle (e.g. *increase up, complete up*). In her study of learner collocations in advanced L2 writing, Nesselhauf (2005) reports another pattern of non-standard MWV use in which a MWV is produced, yet a different MWV form was deemed to be more appropriate. For instance, in the sequence *take over responsibility, take on* rather than *take over* would have been a more acceptable option. Moreover, corpora have shown that L2 speakers are able to combine verbs and particles to form novel MWVs. While the meaning of some of these lexical innovations is difficult to understand or interpret (Gilquin, 2011), other novel MWVs can make perfect sense in the contexts where they are used and might even be formed in analogy with other existing MWVs (Edwards & Laporte, 2015; Gilquin, 2015a; Schneider & Gilquin, 2016). For example, Gilquin (2015a) reported that the MWV *cope up with* as observed in the Singaporean component of the International Corpus of English (ICE) (Greenbaum, 1991) and in the Czech component of the ICLE corpus was probably created in analogy with other PPVs such as *come up with, put up with,* and *meet up with.*

[7] Edwards & Laporte (2015) refer to the case of redundant particles as hyper-explicitness.

Concerning the scope of studies looking at non-standard MWV forms, it should be pointed out that the majority is based on written data from the ICLE and ICE corpora. Also, most of these studies have focused on MWVs containing the particle 'up'. Edwards & Laporte (2015) is an exception to this. Although their study is not entirely devoted to MWV forms, they look at the patterning of 'into' with a few verbs in MWV combinations as used by Dutch speakers of English. Moreover, following Kachru's (1982) model of English varieties[8], most corpus-based research on non-standard MWV forms has adopted a cross-variety perspective, thus comparing MWV use in ESL and EFL varieties. ESL varieties correspond to "institutionalized second-language varieties of English" (Gilquin, 2015a, p. 92) as spoken in countries where English has an official or semi-official language status (e.g. India and Singapur). In contrast, EFL varieties are those used in countries where English is not an official language (e.g. France and Spain). The comparisons of ESL and EFL speakers' production of non-standard MWVs forms have revealed that speakers from both English varieties created novel MWV combinations (Gilquin, 2011; 2015a), which demonstrates that lexical creativity is not an exclusive L1-phenomenon.

### 2.4. MWVs in different L1 verbal systems

L1 background has been frequently investigated as a key factor in L2 use in both second language learning (e.g. Jarvis & Pavlenko, 2008; VanPatten & Williams, 2015) and learner corpus research (e.g. Paquot, 2013; Granger, 2015). Studies looking at L1

[8] Kachru's (1982) model consists of three concentric circles, each circle representing a different English variety. The inner circle includes varieties of English used as a mother tongue (e.g. American English, Australian English, British English). The outer circle represents English varieties used in former British or American colonies where English has the status of a second official language (e.g. Indian English, Singaporean English). The last expanding circle in the model includes EFL varieties usually learned in classroom settings and used in countries where English is not an official language (e.g. Spain, Japan).

background effects on L2 use have often pointed out that L2 speakers tend to rely on their L1 when processing and using the L2 and that they may also be likely to interpret new L2 lexical information with respect to their L1 systems (Sunderman & Kroll, 2006; Schmitt, 2008; Ringbom, 2016). In the context of MWV production, the role of L1 background has been explored, with elicitation and corpus-based studies looking at the frequency of MWVs used by L2 speakers from various L1 backgrounds. In these studies, L1 transfer has often been put forward as the most likely explanation for the patterns of MWV use found in L2 speech. The concept of L1 transfer in MWV research has been explained in terms of whether L2 speakers' L1 is of Germanic or non-Germanic origin (Dagut & Laufer, 1985; Laufer & Eliasson, 1993; Waibel, 2008; González, 2010; Gilquin, 2015b) (see section 2.3.2.1). If the speaker's L1 is Germanic and has MWVs or similar verbal forms, positive L1 transfer has been reported to occur. Therefore, L2 speakers with an L1 of Germanic origin (e.g. Danish and Dutch) may have less difficulty producing MWVs and may be more likely to use MWVs more frequently than speakers with an L1 of non-Germanic origin (e.g. Portuguese and Japanese) in which MWVs are absent.

### 2.4.1. MWVs in the Germanic languages

As mentioned in section 2.2.1, MWVs can be found in a wide variety of Germanic languages including English, proving that they are not "an isolated language-specific phenomenon" (Cappelle, 2014, p. 573). Table 2.4 presents a list of the present-day Germanic languages in addition to English in which MWVs have been observed along with examples of MWVs in each of those languages. Particles are italicised in the table.

Table 2.4 MWV use in the Germanic languages

| **Germanic language** | **Example(s)** | **Reference** |
|---|---|---|
| German | *auf*stehen<br>(get up) | Lüdeling, Hirschmann & Shadrova (2017) |
| Dutch | *neer*stortte<br>(fall down, past tense) | Booij (2002) |
| Yiddish | *oj*strinkən<br>(drink up) | Dehé (2015) |
| Norwegian | slå *opp*<br>(look up) | Dehé (2015) |
| Swedish | ringa *upp*<br>(ring up) | Dehé (2015) |
| Danish | malede *over*<br>(paint over, past tense) | Braunmüller (1999) as cited in Thim (2012) |
| Icelandic | standa *upp*<br>(stand/get up) | Dehé (2015) |
| Faroese | sláa *upp*<br>(look up) | Dehé (2015) |
| West Frisian | siket *troch*<br>(search through) | Hoekstra (2001) |

According to Dehé (2015), particles in German, Dutch and Yiddish MWVs precede and attach to verbal stems in infinitive forms such as the German verb *aufgeben* (i.e. *to give up*, *auf* meaning '*up*'), but they can split from the verb in the presence of a direct object noun phrase (e.g. *Sie gab ihre Arbeit auf*, meaning *'She gave up her job'*) (Cappelle, 2014). In contrast, Present-day English, Norwegian, Swedish, Danish, Icelandic, Faroese, and West Frisian MWVs contain particles taking the post verbal position in infinitive forms (Thim, 2012). Close parallels to the Germanic MWV have also been reported in languages that belong to more genetically distant families such as Old

French (Dufresne, Dupuis & Tremblay, 2003) and Hungarian (Ladányi, 2015). It is beyond the scope of the present study to provide a full account of the morphosyntactic and semantic properties of MWVs in all Germanic and non-Germanic languages where they have been observed. However, the following sections are devoted to the description of MWVs and similar verbal forms in Chinese, Italian, and Spanish – the three L1 backgrounds in which MWV use is analysed in the present study.

### 2.4.2. Chinese

Chinese possesses a group of verbal structures that broadly resemble English MWVs in the way they encode motion events. These verbal forms have been referred to as directional verbal compounds (Chen, 2008) and multi-morpheme motion constructions (Lin, 2015). The internal structure of a directional verbal compound is morphologically more complex than that of the English MWV. Typically, directional verbal compounds have two or even three verbs that appear together as one unit (e.g. *xia-qu*, *go down; reng-chu-qu*, *throw out*) and whose order in the compound cannot be reversed (Lin, 2015). While the first element in forms like the previous ones is often considered the main verb (e.g. *xia* = descend, *reng* = throw), the second verb is treated as the element indicating direction (Talmy, 1985, 2000). In the examples above, *qu* and *chu* translate as *go* and *exit* when they appear alone, but in verbal compounds like *xiaqu* and *rengchuqu* their meaning is equivalent to that of the particles *down* and *out* respectively. As opposed to MWVs in the Germanic languages and in the Romance languages (see sections 2.4.3 and 2.4.4), all the components of the Chinese verbal compounds can act as independent verbs (Slobin, 2004). This means that even those verbs that function as particles can appear alone and have a full verb status.

In terms of semantics, the so-called Chinese particle verbs are more restricted than those in English. Chinese verbal compounds tend to express mainly literal meanings and the particles in them indicate direction. Nonetheless, Chen (2008) argues that instances of metaphorical verbal compounds are not uncommon. For example, the verbal compound *shuo-chu-lai* has three elements which literally mean *speak-exit-come*. The meaning of such verb form, however, corresponds to that of the MWV *speak out* (Chen, 2008). Similarly, the individual meanings of the elements in the compound *cai-chu-lai* are *guess-exit-come*, yet the meaning of the compound does not correspond to the sum of the individual meaning of each element but is equivalent to that of the MWV *figure out*.

### 2.4.3. Italian

Contrary to other Romance languages, Italian makes relatively extensive use of MWVs, with several verb + particle forms being particularly common in speech (Masini, 2005). The productivity of Italian MWV forms and their attested uses in contemporary Italian possibly explain the thorough treatment that they have received in the literature (e.g. Iacobini, 2006, 2009a, 2009b; Iacobini & Masini, 2006; Maiden & Robustelli, 2014). In terms of form, Italian MWVs resemble the English ones. They consist of two elements: a verb and an adverbial particle that takes post verbal position (e.g. *l'ascensore va su, the elevator goes up*) (Iacobini, 2015). The verb and the particle display a strong sytantic and semantic bond, which is mainly demonstrated by the fixedness of the two components (Iacobini, 2009a). Unlike the English MWV form, the verb and the particle in Italian MWVs often do not welcome intervening elements with the exception of clitics attaching to the verbal base (e.g. *Devi buttarla fuori, You must throw her out*) and some adverbs (e.g. *Non devi andare mai indietro, You must go never back*) (Iacobini, 2015). When the referent in a sentence containing a MWV is realised

in the form of a nominal or pronominal form rather than as a clitic, such referent has to be preceded by the preposition *a* (i.e. to) as the following examples show:

| | |
|---|---|
| -Referent is expressed by a clitic: | *saltargli addosso*<br>*to jump him on* (i.e. *to jump on him*) |
| -Referent is expressed by a nominal form: | *saltare addosso* ***al*** *nemico*<br>*to jump on* ***to*** *the enemy* (i.e. *to jump on the enemy*) |

The unity between the verb and the particle in Italian MWVs is further evidenced in their transitivity patterns. Italian MWVs can be both transitive and intransitive, with direct objects immediately following the particles rather than occurring in between the two elements. However, Masini (2008) explains that, in some cases and for the sake of emphasis, nouns and noun phrases can occur in between the verb and the particle (e.g. *mettere la palla dentro, to put the ball in*). These cases occur more frequently in spoken than in written registers though.

Semantically, MWVs in Italian also exhibit different degrees of compositionality. Although most MWVs convey literal and directional meanings (e.g. *andare via, go away*), there are instances of aspectual and idiomatic MWVs. According to Iacobini (2015), a characteristic of MWVs that distinguishes Italian from other Romance languages where MWVs have been observed is the presence of aspectual verb + particle sequences, i.e. verbs followed by particles that indicate completion or result of the action expressed by the verbs (e.g. *strofinare via le macchie, to rub off the stains*). MWVs that require a metaphorical interpretation are not uncommon in Italian either.

For instance, the MWV *tirare avanti* literally translates as *pull forward*. However, in examples like *un po' di soldi per poter tirare avanti* (i.e. *a little money in order to survive*), the MWV means *to get by, survive or resist* (Iacobini, 2015). Another example of an idiomatic MWV is *fare fuori*. It literally translates as *do out* (do=fare, out=fuori), yet its meaning is that of the verb *kill*.

### 2.4.4. Spanish

According to Mateu & Rigau (2010), MWVs were commonly used in all Romance languages, not only in Italian, until the mid-15$^{th}$ century. However, MWVs in Spanish as well as in other Romance languages like French went through a marginalisation process which, together with the emergence and more frequent use of path verbs (i.e. verbs expressing direction or manner of motion such as *exit*), resulted in the decrease and almost complete loss of MWVs.

Unlike in Italian, Spanish MWVs are not a productive class. In fact, Iacobini (2015) argues that Spanish MWVs comprise an extremely limited set of no more than fifty items. With respect to their form, MWVs in Spanish also consist of two elements: a verb and a particle. The particle is mainly of motion, and it indicates the direction of the action expressed by the verb (e.g. *salir <u>adelante</u>, to go <u>on</u>; volver <u>atrás</u>, to turn <u>back</u>*) (Hijazo-Gascón & Ibarretxe-Antuñano, 2013). It is possible to find MWVs in which a verb base that is already indicating the path (i.e. the direction of the action) (e.g. *salir, exit*) is followed by a particle also expressing path and thus reinforcing the expression of direction (e.g. *salir afuera, exit out; meter adentro, push inside; subir arriba, ascend up*). These MWVs in which both the verb and the particle seem to express direction, however, tend to be more common in spoken registers and might even be considered

errors or non-standard uses in Spanish (González Fernández, 1997). Syntactically, MWVs in Spanish can display different transitivity patterns and even take intervening elements between the verb and the adverbial particle. Semantically, unlike English MWVs, most Spanish MWVs are fully compositional and literal in meaning, with a limited number of forms used idiomatically (Iacobini, 2015) as the examples in Table 2.5 show.

Table 2.5 Idiomatic MWVs in Spanish

| **MWV** | **Literal meaning** | **Idiomatic meaning** |
|---|---|---|
| *echarse atrás* | go back | back out |
| *sacar adelante* | get forward | get ahead |
| *venirse abajo* | come down | break down |

As shown in the previous descriptions of MWVs and similar verbal forms in Chinese, Italian and Spanish, the three L1 backgrounds represent different linguistic patterns of MWV use. The differences are mainly related to three aspects: the frequency of MWV occurrence, form of MWVs and degrees of compositionality. As opposed to MWV-like structures in Chinese and Spanish, MWVs in Italian represent a rather productive class and have been found to occur very frequently in spoken discourse. In terms of form, MWV-like structures in Chinese differ considerably from those found in the two Romance languages (i.e. Italian and Spanish) and in English given that they consist of combinations of two or three verbs that appear together and function as one unit. Finally, while MWVs in Chinese and Spanish tend to be, mainly, fully compositional and contain particles expressing direction of motion, Italian MWVs can display the full range of compositionality types found in English MWVs.

## 2.5. The L2 production of MWVs in different task types

Research has demonstrated that task type exerts great influence on the language that L2 speakers produce and on the ways in which they produce it (Foster & Skehan, 1996; Newton & Kennedy, 1996; Hinkel, 2009). In fact, the various requirements of specific tasks in terms of cognitive demands, topics, grammar, vocabulary, and communicative goals tend to result in different types of L2 production. In learner corpus research, the effect of task type on L2 use has been relatively underexplored (Alexopoulou et al., 2017; Caines & Buttery, 2018), and the same holds true for task effects on MWV production. Little research has been conducted that focuses directly on the effect of different tasks on the type and frequency of MWVs produced by L2 speakers. Most findings in this respect come from studies using tests to elicit specific target MWVs and assessing L2 speakers' knowledge of these verbs rather than their use in naturally occurring discourse. Because of the research approach in these studies, the terms 'tests' and 'tasks' have tended to be used synonymously. While the differences between a test and a task are well documented (Samuda & Bygate, 2008) and taken into consideration in the present study, for the sake of clarity, the term task will be used when discussing both.

### 2.5.1. Overview of task types used in the study of MWVs

The tasks in which L2 speakers' use of MWVs has been examined broadly fall into three categories: reproduction, recognition, and production tasks. The first category corresponds to tasks in which the L2 speakers have to reproduce target MWVs, for example, by translating them from their L1 and into the L2. Recognition tasks typically require L2 speakers to identify target MWVs from a set of options (e.g. multiple choice)

whereas production tasks involve using MWVs in writing or speaking (e.g. free writing and story re-tell). Table 2.6 provides an overview of the task types used to evaluate L2 production of MWVs together with a description of the task format and reference studies whose research design included those tasks. As can be seen in the table, the majority of tasks aimed to evaluate MWVs in writing rather than in speaking. The number of MWVs analysed in each task usually ranged from ten to twenty.

Table 2.6 Overview of task types where L2 MWV use has been evaluated

| Task type | Output type | Description of task format | Reference studies |
|---|---|---|---|
| **Multiple choice** | written | -Sentences or dialogues in English with blanks to be filled with one of four alternatives provided.<br><br>-The alternatives include: the appropriate MWV, an appropriate and semantically equivalent one-word verb, and two distractors (i.e. another MWV and a non-phrasal single-word verb). | Dagut & Laufer (1985), Hulstijn & Marchena (1989), Laufer & Eliasson (1993), Sjöholm (1995), You (1999), Liao & Fukuya (2004), Saiya (2011), Kamarudin (2013), Sara & Mohammadreza (2013), Barekat & Baniasady (2014), Becker (2014) |
| | | -A variation of the task can include only two alternatives: a MWV and a semantically equivalent one-word verb. | Kharitonova (2013) |

| | | | |
|---|---|---|---|
| **Translation** | written | -Sentences or dialogues in English with MWVs left out. L1 translations of the missing MWVs are provided. Speakers have to translate the L1 versions into English. | Dagut & Laufer (1985), Hulstijn & Marchena (1989), Irujo (1993), Laufer & Eliasson (1993), You (1999), Liao & Fukuya (2004), Ayadi (2010), Sara & Mohammadreza (2013), Barekat & Baniasady (2014), El-Dakhs (2016) |
| | | -In a variation of the task, sentences or dialogues in the speakers' L1 are provided. The MWVs to be translated into English are underlined. | Becker (2014) |
| **Memorisation / re-call** | written | -English sentences containing MWVs and their L1 translations are provided. Speakers have ten minutes to memorise the sentences. After one hour, speakers are given the same English sentences with the MWVs left out. They need to supply the missing MWV forms. | Dagut & Laufer (1985); Barekat & Baniasady (2014) |

| | | | |
|---|---|---|---|
| | | -In a variation of the task, the L1 translations of the sentences containing the MWVs are not provided. | Hulstijn & Marchena (1989), Liao & Fukuya (2004) |
| **Verb elicitation** | written | -Incomplete sentences in English are provided. Speakers should complete the sentences by supplying an appropriate MWV form. Pictures depicting the actions expressed in the sentences are provided as well. | You (1999) |
| **Cloze** | written | -Dialogues in English contain missing parts which correspond to target MWVs. The task consists of restoring the missing parts. | Nassaji & Tian (2010), Karimi (2017) |
| **Editing** | written | -Dialogues in English contain erroneous sections related to the use of target MWVs. The task consists of identifying and correcting the erroneous parts. | Nassaji & Tian (2010), Karimi (2017) |
| **Free writing** | written | -Learners write a narrative composition on a topic assigned by the researcher. Target MWVs have to be included in the composition. | Barekat & Baniasady (2014) |

| | | | |
|---|---|---|---|
| | | -In a variation of the task, compositions are written based on topics selected by the speakers. | Mahmoud (2015) |
| **Story re-tell** | spoken | -A story containing the target MWVs is given to the speakers in their L1. Learners can read the story multiple times for a few minutes and are then asked to re-tell as much as they remember in English. | Becker (2014) |

### 2.5.2. The influence of task type in L2 production of MWVs

The evidence of the effects of task type on L2 MWV production mainly comes from elicitation studies. Building on Dagut & Laufer's (1985) research design, most elicitation studies have focused on the knowledge of selected MWV forms that L2 speakers have as elicited and assessed by means of different controlled tasks (see sections 2.5 and 2.5.1). More specifically, these studies have looked at the interaction between task type and i) the frequency of use of the selected MWVs, ii) lexical verb type (i.e. MWVs vs. one-word verbs), and iii) semantic category of MWV (i.e. figurative vs. literal).

Findings from those studies evaluating MWV production in different tasks appear so far to be inconclusive. While Sara & Mohammadreza (2013) found no relationship between task type and MWV use, other authors have reported that translation tasks lead to greater avoidance of figurative MWVs (Laufer & Elliason, 1993; Liao & Fukuya, 2004; Becker, 2014). Translation tasks have also been found to display a higher frequency of one-word verbs compared to MWVs (El-Dakhs, 2016) although Barekat & Baniasady (2014) report different results. Moreover, findings are also mixed when describing the effect of recognition tasks (e.g. multiple choice) on L2 production of MWVs. While Laufer & Eliasson (1993) show that multiple choice tasks are the ones most likely to yield greater avoidance of literal MWVs, Becker (2014) and You (1999) found not only that literal MWVs are preferred over figurative ones in multiple choice tasks but also that speakers used more MWVs overall in this particular task.

Considering the tasks that have been used to investigate MWVs, the story re-tell is worthy of attention as it appears to be the only speaking task assessing the use of MWVs

in oral discourse. In a study conducted by Becker (2014), the story re-tell task had the lowest frequency of MWVs compared to the other two tasks he investigated (i.e. translation and multiple choice). This finding was attributed to the demands of the task. Unlike the translation and multiple-choice tasks, the story re-tell required L2 speakers to first understand the meaning of the MWVs to be able to later use them meaningfully in speaking, which was considered more cognitively demanding than mere recognition of target MWV forms. Surprisingly, despite the low MWV frequency in the story re-tell task, speakers demonstrated a preference for figurative MWVs compared to literal ones. Becker explained this last finding in terms of attention as he argues that it was likely that the presence of several figurative MWVs in a relatively short story might have drawn L2 speakers' attention.

Although often the design of the tasks and the target MWVs have been broadly similar across studies, the reasons why findings about the influence of task type on MWV use have tended to be inconsistent might partly be due to two factors. The first factor corresponds to the differences in the definition of the MWV avoidance construct (e.g. Liao & Fukuya, 2004; El-Dakhs, 2016). The second factor relates to the ways in which researchers controlled (or not) for knowledge of the target MWV forms. It is also worth mentioning that the design of some tasks appeared to have strongly encouraged the use of MWVs, for example, by providing only two alternatives (a MWV and a one-word verb) in a multiple choice task (Kharitonova, 2013). Finally, as can be seen in Table 2.6, little is known about the link between spoken tasks, let alone spoken tasks in learner corpora, and L2 production of MWVs. Further research is needed to examine the extent to which L2 speakers' use of MWVs is affected by the speaking tasks in which they

engage. One of the aims of the present study is to contribute to bridging this gap in knowledge.

### 2.6. Research questions

This study seeks to investigate how L2 speakers use MWVs in spoken, interactive communication. To this end, it looks at MWVs from three different perspectives: i) the MWVs themselves and their usage patterns in the TLC, ii) the L2 speakers who produce those MWVs, and iii) the tasks in which the MWVs are used. Considering those three perspectives, the study aims to answer the following research questions:

RQ1: How are MWVs, i.e. PVs and PPVs, used by L2 speakers in terms of i) MWV frequency and coverage, ii) lexical verb and particle productivity, iii) number and type of non-canonical MWV forms, and iv) polysemy?

RQ2: What is the effect of learner variables, i.e. L2 proficiency and L1 background, on L2 speakers' use of MWVs?

RQ3: What is the effect of text-specific variables, i.e. task type, on L2 speakers' use of MWVs?

RQ1 looks at MWVs from the verb perspective. Its main aim is to provide a comprehensive description of L2 speakers' use of MWVs not only by reporting the frequency of MWV occurrence as it has been done in earlier corpus-based work on the topic but also by revealing patterns of non-standard MWV use and providing an account of L2 speakers' semantic MWV knowledge. Let us know operationalise RQ1 by

defining the four aspects it involves. First, MWV coverage is defined as the percentage of MWV occurrences that each MWV lemma accounts for. Second, productivity refers to the ability of the verbal element or particle in a MWV to combine with other particles or verbs respectively in order to form lexically distinct MWVs. Third, non-canonical forms are understood as combinations of lexical verbs and particles that are not attested in the MWV dictionary consulted for the manual analysis of data (see sections 3.3 and 3.10.3). Finally, polysemy is defined as the number and frequency of meaning senses associated with the most frequent MWVs.

RQ2 analyses MWVs from the L2 speakers' perspective. In particular, it takes into consideration two of the most prominent variables that are likely to affect the acquisition and use of vocabulary in L2 contexts: level of L2 proficiency and L1 background (Jarvis, 2000; Nation, 2001; Ringbom, 2016; Conklin & Carrol, 2019). On the one hand, L2 proficiency was operationalised as L2 speakers' ability to use English in a spoken context as measured by the Graded Examinations in Spoken English (GESE) (see section 3.4). Three broad proficiency bands according the Common European Framework of Reference (CEFR) are used in the study: B1 (threshold speakers), B2 (intermediate speakers) and C1-C2 (advanced speakers). On the other hand, L1 background was understood as L2 speakers' native language as reported in the corpus metadata (see section 3.4). The following L1s are used: Chinese, Italian and Spanish. For research questions 2 and 3, 'MWV use' was operationalised in terms of two inter-related linguistic variables: the frequency of MWV occurrence and the range of MWVs.

The aim of RQ3 is to analyse MWVs from the perspective of tasks. More specifically, RQ3 investigates the extent to which L2 speakers' use of MWVs varies depending on the type of speaking tasks in which they engage. In the study, the speaking tasks correspond to the different free-production activities that L2 speakers had to carry out as part of their examination of spoken English (i.e. GESE) and include: presentation, discussion, interactive task, and conversation. The format and discourse characteristics of these tasks are described in section 3.5.3.

## 2.7. Summary

This chapter focused on four main areas: i) the description of MWVs in English, ii) the ways in which MWVs have been approached in L1 and L2 corpus linguistic studies, iii) an overview of MWVs in Chinese, Italian and Spanish, and iv) the relationship between different task types and L2 production of MWVs. With respect to the first area, the chapter explained that MWVs are verbal forms characteristic of the Germanic languages rather than particularly English. The Present-day English MWV originated from a series of linguistic processes that first led to a separation of verbs and particles, which were fused together in Old English, to the post position of particles that characterises MWVs today.

The chapter then proceeded to define the two types of MWVs that are the object of investigation of the present study: PVs and PPVs. Special emphasis was given to differentiating PVs from single lexical units despite morphological and neurolinguistic approaches arguing for the word-like status of PVs. Differences between PVs and two types of MWUs, namely grammatical collocations and idioms, were also reviewed and explained. While transparent PVs have been considered collocations, non-transparent

PVs tend to be defined as idioms. Arguments were provided in favour of treating PVs and PPVs as MWUs rather than as collocations or idioms given that those last two terms tend to focus mainly on the semantics of MWVs and disregard their syntactic properties. Next, the case of 'free combinations' (i.e. verb + particle sequences with literal, transparent meanings) was further explored. Comprehensive grammars (e.g. Quirk et al., 1985; Biber et al., 1999) have excluded literal verb + particle combinations from the MWV group, a decision not followed in this study given that it is difficult to semantically categorise MWVs and that literal verb + particle combinations also exhibit the same syntactic properties of other semi- and non-compositional MWVs.

The chapter went on to report the main contributions of corpus linguistics to the study of MWVs. From the perspective of L1 corpora, research has not only attested the high frequency of MWVs in English, but it has also revealed that the meaning senses associated with a single, polysemous MWV are not equally frequent. In terms of MWV semantics, historical corpora have been particularly valuable in demonstrating the origin and evolution of meaning senses and particle meanings. L1 corpora representing different registers have also served to describe the distribution of MWVs in written and spoken discourse, the latter being the most common environment for MWVs to occur. With respect to L2 corpora, contributions have gone beyond the mere identification of MWV frequencies and have thrown light on areas such as the role of L2 speakers' L1 in MWV production, the distribution of MWVs particularly in written registers, and the uses of non-compositional, semi-compositional and transparent MWVs. L2 corpora have also proven useful in describing speakers' use of non-standard MWVs by showing a variety of patterns that range from the addition of a redundant particle and particle misplacement to the creation of novel MWV forms.

In the last sections of the chapter, an overview of MWVs and similar verbal forms in Chinese, Italian and Spanish was provided. On the one hand, Italian and Spanish MWVs consist of verbs followed by adverbial particles, thus broadly resembling English MWVs. MWVs in Italian, however, are part of a much more productive category than those found in Spanish. On the other hand, directional verbal compounds in Chinese have been found to be approximate equivalents to the English MWV in the sense that their elements carry verb + particle information when used together in the compounds.

Finally, the influence of task type in the L2 production of MWVs was explored. Reproductive, reception and mainly written production tasks have been used in elicitation studies to investigate (figurative) MWV avoidance in L2 contexts. Research findings from these studies indicate that the effect of task type on MWV use remains unclear. It is argued that the mixed results from earlier studies might have been due to the various ways in which avoidance was operationalised and the possible lack of knowledge of the target MWVs by participants in these studies. The chapter finished by outlining and operationalising the three research questions that the study aims to answer.

# Chapter 3: Methodology

## 3.1. Introduction

This chapter describes the research approach and methodological decisions made in this study. First, it provides an overview of the research design followed by an explanation of the main terminology employed in this thesis. Second, it introduces the corpus used to carry out the analyses. Details regarding the data collection procedure including the identification and extraction of MWVs follow as well as a description of the manual coding of the data and evaluation of polysemy. The chapter concludes by discussing the statistical analyses used to meet the needs of the research and the rationale behind their selection.

## 3.2. Research design

This study draws on second language acquisition theory and corpus linguistics to inform language analysis. It adopts a descriptive corpus-based approach (Tognini-Bonelli, 2001; McEnery, Xiao & Tono, 2006). As such, corpus data was used to primarily explore L2 speaker language in light of the existing linguistic theory on MWVs (see chapter 2). Following Granger (1998), the data was examined using the hypothesis-finding approach given that no hypothesis of L2 MWV use was put forward prior to the analysis of the data. The study can also be described as being comparative in nature (see section 1.3.1). Rather than comparing L2 production against a control corpus of L1 English, it is based on the comparison of different L2 English varieties in terms of L2 proficiency and L1 background and across task types. A comparison with an L1 English corpus was not pursued for two reasons. First, the aim of the study is to determine

whether the observed patterns occur in the language produced by a specific group of L2 speakers. Second, an L1 corpus that is comparable to the TLC was not available at the time of this study. A further methodological issue relates to the quasi-longitudinal design that this study follows in that the corpus data examined was gathered at a single point in time but comes from several groups of L2 speakers at various levels of L2 proficiency (Huat, 2012; Callies, 2015; Meunier, 2015).

The study makes use of a combination of both qualitative and quantitative analyses. The first ones allowed an in-depth exploration of MWVs in their related linguistic environment and the description of patterns associated with the use of MWVs in context. The second ones allowed the quantification of the data including frequency counts of MWVs according to L2 proficiency level, L1 background, and task type. Both descriptive and inferential statistics were used to account for inter-speaker variation. Further details of the statistical procedures are provided in section 3.12.

### 3.3. Terminology adopted in the study

This section explains the terminology that is used most extensively throughout this thesis. The terms described below mainly refer to the types of verb forms analysed in this study. The section also provides reasons as to why certain terms that are commonly used to refer to MWVs were not adopted in this study. The use of other key terms (i.e. L2 proficiency, L1 background and speaking task) is discussed in section 2.6 as part of the operationalisation of the research questions.

Two types of MWVs are analysed in this investigation: PVs (e.g. *take out*) and PPVs (e.g. *look forward to*) (see sections 2.2.2 and 2.2.3). Unless otherwise specified, the

term MWV is used to refer to these two types of verbs only. A special mention should be made regarding the term 'phrasal verb' (PV), which is favoured over other less common terms used in the literature to refer to such verb type (e.g. particle verbs, verb-particle constructions, separable verbs, compound verbs, poly-word verbs, and discontinuous verbs). In particular, the term 'verb-particle construction' is not used to avoid confusion with other more complex and longer sequences involving PVs and noun phrases in agent and patient positions such as *he dropped the children off* (e.g. Baldwin & Villavicencio, 2002; Gries, 2003). Also, calling PVs constructions might presuppose commitment to Construction Grammar (Goldberg, 2003; Hoffmann & Trousdale, 2013), whose tenets are different from those on which the present study is based. Occasionally, however, I mention the terms 'particle verb' and 'verb-particle combination' to refer to PVs specifically when I quote authors who have used this term in their work (e.g. Sroka, 1972).

Another key term in this study is 'non-canonical MWV'. Non-canonical MWV forms correspond to combinations of a lexical verb and an adverbial particle that are not documented in the dictionary used for the manual analysis of data (see section 3.10.3.3). It is important to notice that the term 'non-standard MWV forms' is used synonymously with 'non-canonical MWVs'. The label 'error' is particularly avoided to refer to non-canonical MWVs as it might imply that the L2 data were subject to an error analysis of the kind described by Corder (1981), which was not carried out in the present study. Moreover, the label 'error' is avoided as there is no a priori assumption that all non-canonical MWVs are instances of erroneous use.

### 3.4. The Trinity Lancaster Corpus

The Trinity Lancaster Corpus (TLC) (Gablasova, Brezina & McEnery, forthcoming) is a large-scale, POS-tagged corpus of L1-L2 speech consisting of 4.2 million running words. It is the result of a collaborative research project between the ESRC Centre for Corpus Approaches to Social Science (CASS), Lancaster University, and Trinity College London, a major international examination board. The corpus was created from recordings of the Graded Examinations in Spoken English (GESE) conducted by Trinity College London. GESE is a twelve-grade examination assessing spoken English proficiency from beginning to advanced levels; each GESE grade in turn is mapped to one of the Common European Framework of Reference (CEFR) levels. From the total of twelve grades, interactions from six grades were used to create the corpus as shown in Table 3.1. GESE examinations are designed for speakers of languages other than English, and their primary objective is to provide measures of spoken linguistic competence based on candidates' performance in a range of speaking tasks (Trinity College London, 2019). Trained examiners who are native speakers of English are in charge of conducting the examination and evaluating speakers' performance.

Table 3.1 GESE grades used in the TLC

| **GESE grade** | **Stage of L2 development** | **CEFR level** |
|---|---|---|
| **6** | Threshold | B1.2 |
| **7** | Intermediate | B2.1 |
| **8** | Intermediate | B2.2 |
| **10** | Advanced | C1.1 |
| **11** | Advanced | C1.2 |
| **12** | Advanced | C2 |

## 3.5. Corpus design

### 3.5.1. Learner variables

The data used in this study come from a total of 1,927 exam candidates (L2 speakers) whose speech contain 2,446,246 tokens in total. These L2 data were produced by L2 speakers across four proficiency bands according to the CEFR (i.e. B1, B2, C1, and C2). The corpus also contains more specific information about speakers' overall performance within each proficiency band as it indicates the exam achievement mark (A= distinction, B= merit, C= pass) and individual mark that speakers received in every task (A= excellent, B= good, C= satisfactory, D= unsatisfactory) (Trinity College London, 2018). The overall size and proportions for each proficiency band as well as the distribution of speakers per overall achievement in the exam are shown in Table 3.2.

Table 3.2 Distribution of L2 speakers and tokens per proficiency band in the TLC

| CEFR band | Number of speakers | Exam achievement mark | Number of tokens | Percentage of tokens in the corpus |
|---|---|---|---|---|
| **B1** | 863 | Distinction: 279<br>Merit: 372<br>Pass: 212 | 711,598 | 29.1% |
| **B2** | 756 | Distinction: 136<br>Merit: 367<br>Pass: 253 | 1,026,237 | 41.9% |
| **C1** | 240 | Distinction: 44<br>Merit: 71<br>Pass: 125 | 544,157 | 22.2% |
| **C2** | 68 | Distinction: 5<br>Merit: 24<br>Pass: 39 | 164,254 | 6.7% |
| Total | 1927 | 1927 | 2,446,246 | 100% |

As far as the age of the L2 speakers is concerned, the data come from speakers whose ages range from 8 to 72 years old. The speakers are divided into five main age bands according to the ages that they reported at the time that the examination took place: young speakers (8-15), adolescents (16-19), young adults (20-35), middle adults (36-50), and older adults (51 and older). Candidates' ages at which they were first exposed to English also vary and range from birth to 60 years old (Gablasova, 2017). Regarding speakers' previous learning history, the data were collected about speakers' highest level of completed education (from primary to tertiary education), uses of English (at school, at work, with family, with friends, on the Internet), and pattern of acquisition (in a foreign language classroom, in English-medium subject classes, in an English-speaking country, from TV or the Internet, and through self-study).

With respect to speakers' L1s, the speech samples come from a vast variety of cultural (e.g. Mexico, Argentina, Brazil, Portugal, and Sri Lanka) and linguistic backgrounds (e.g. Spanish, Portuguese, Sinhala, and Tamil). Despite this variety, there is a higher proportion of some L1 groups in the corpus (i.e. Chinese, Italian, Russian, and Spanish) (Gablasova, Brezina & McEnery, forthcoming). In addition, the corpus contains data about L2 speakers who reported having more than L1 background (e.g. Italian-German, Spanish-English, Marathi-Gujarati, and Chinese-Cantonese).

### 3.5.2. Examiner-related variables

The examiner data come from 140 examiners, who produced a total of 1,829,495 tokens. They represent three different age groups: young adults (20 to 35 years old), middle adults (36-50 years old) and older adults (51 years old and over) (Gablasova, 2017). As mentioned previously, examiners are all native speakers of English and have received rigorous training in the assessment of L2 speakers' performance in the GESE examinations.

### 3.5.3. Speaking tasks

The structure of the TLC is built upon L2 speakers' performance in a series of elicited monologic and dialogic tasks that increase in number depending on the grade at which speakers are assessed (Trinity College London, 2018). Most tasks are semi-formal and last approximately five minutes each. At grade 6, the format of the examination consists of two tasks: discussion and conversation. At grades 7 and 8, speakers engage in three tasks: discussion, interactive task and conversation. At grades 10, 11 and 12, the most advanced levels, speakers are assessed in five tasks: presentation, discussion,

interactive task, listening task, and conversation. Table 3.3 summarises these tasks and shows the interaction type and level of formality of each.

In the *discussion* task, candidates use a self-selected topic to engage in an exchange of ideas with the examiner, who asks questions about this topic based on any notes that the candidate provides or the examiner's own notes made during the task. It is also candidates' responsibility to ask examiners questions during the discussion. In the most advanced levels (10 to 12), the discussion is based on the ideas raised by candidates in a previous presentation task. *Conversation* requires speakers to talk with examiners about two subject areas which are selected by the examiner and vary according to the level. In grades 6 to 11, subject areas come from a list of topics which the candidate is familiar with (e.g. money, fashion, village and city life, use of the Internet) whereas in grade 12 the examiner can select any topic that is considered appropriate. In the *interactive task*, examiners provide candidates with an oral prompt which they should use to find out more information, ask examiners questions and comment on the examiner's responses. The following is an example of a prompt used in the interactive task: *Many people think that we should spend more time enjoying ourselves and less time trying to make money, but I'm not sure you can do one without the other* (Trinity College London, 2019). In the *presentation* task, candidates deliver a formal presentation about a self-selected topic. This task is mainly monologic; there is little interaction between examiners and candidates during this phase of the examination. Examiners, however, take notes during the presentation, which they will later use to facilitate discussion of the topic once the presentation finishes. Finally, *listening* is subdivided into two types of listening tasks. First, candidates are orally provided with two short prompts to which they should respond by suggesting a suitable ending.

Second, candidates receive an additional prompt and are required to identify contexts and participants. The following is an example of a prompt used in the first listening task: *I don't like being told what to do. If I buy something that I need to assemble, I never look at the instructions. When I buy electronic gadgets I just fiddle around until they start working. And when I'm getting a meal ready I just throw all the ingredients together and rely on my taste instead of using a…* (Trinity College London, 2019). As the sample prompt shows, expected answers may vary and many different ways to finish the passage are possible (e.g. recipe, cookbook, etc.).

Table 3.3 Overview of tasks in the TLC

| **Grade** | **Task** | **Type of interaction** | **Level of formality** |
|---|---|---|---|
| **6** | discussion | jointly-led | semi-formal |
| | conversation | jointly-led | semi-formal |
| **7, 8** | discussion | jointly-led | semi-formal |
| | interactive task | candidate-led | semi-formal |
| | conversation | jointly-led | semi-formal |
| **10, 11, 12** | presentation | candidate-led | (semi-) formal |
| | discussion | jointly-led | semi-formal |
| | interactive task | candidate-led | semi-formal |
| | listening task | examiner-led | semi-formal |
| | conversation | jointly-led | semi-formal |

### 3.6. Corpus selection

The TLC is well suited for the study of MWVs in L2 communication. Four specific criteria motivated its selection: medium of communication, L1 backgrounds, proficiency measures, and task variability. First, while there has been a tendency to explore MWVs, particularly PVs, in written corpora (see section 2.3.2.3), the TLC provides a relatively unique opportunity to investigate MWVs in L2 spoken communication, which constitutes a more common environment for MWVs to occur (Biber et al., 1999; Gardner & Davies, 2007) (see section 2.3.1.3). Second, one of the aims of this study was to investigate L2 speakers' use of MWVs across L1 backgrounds; therefore, the corpus data had to allow the comparison of different L1 populations. Being a multi-lingual corpus, the TLC contains information about a wide variety of L1 backgrounds. Such information is often not clearly indicated in other learner corpora, which might only report speakers' nationalities in their metadata (Gilquin, 2015b). Third, the TLC was also selected because it provides explicit and reliable information about L2 speakers' proficiency level, an essential variable in this study. Given the various ways in which L2 proficiency has been operationalised in learner corpus studies and the lack of rigour when documenting proficiency in corpus metadata (Callies, 2015), it was important to select a corpus which clearly recorded L2 proficiency. Also, as Carlsen (2012) suggests, the corpus should reliably link the assignment of proficiency levels to a solid assessment scheme such as the CEFR, thus allowing the analysis of MWVs across L2 populations at different stages of language development. As was described in section 3.5.1, the TLC provides information about speakers' L2 proficiency not only at the grade level (e.g. grade 12 = C2) but also at the task level. Finally, the TLC was also selected because it allows the analysis of MWVs as produced by L2 speakers across a range of speaking tasks in contrast to other available corpora

where L2 data were elicited in one or two tasks only. Since not all tasks in a learner corpus afford equal opportunities to observe the linguistic variables of interest (Tracy-Ventura & Myles, 2015), task variability was a priority when selecting the corpus especially because another aim of this study was to determine the effect of task type on L2 speakers' use of MWVs.

### 3.7. Sub-corpora used in the study

Based on the research aims of this investigation, different components of the TLC were used to answer the research questions. These sub-corpora only contain L2 speaker language and are presented in Table 3.4.

Table 3.4 Structure and size of sub-corpora used to answer research questions

| L2 proficiency | L1 backgrounds | Number of speakers | Number of tokens |
|---|---|---|---|
| **B1** | Chinese | 108 | 87,353 |
| (Grade 6) | Italian | 149 | 117,572 |
| | Spanish | 349 | 265,195 |
| **B2** | Chinese | 96 | 117,385 |
| (Grades 7 and 8) | Italian | 129 | 172,974 |
| | Spanish | 277 | 360,332 |
| **C1-C2** | Chinese | 33 | 73,001 |
| (Grades 10 to 12) | Italian | 68 | 157,945 |
| | Spanish | 139 | 301,434 |

As can be seen in the table, C1 and C2 proficiency bands were grouped together and represent advanced L2 production. This decision was made based on the overall low number of C2 speakers from the L1 backgrounds selected (i.e. 41). In addition, despite

the variety of L1 backgrounds available, the study draws on data from Chinese, Italian and Spanish speakers only. As mentioned in section 3.5.1, those three L1 groups have the highest number of speakers in the TLC and account for 70% of all L2 speakers in the corpus. It is worth noting that only speakers who reported one L1 background (Chinese, Italian or Spanish) were included in the study to control for the possible effect of knowledge of another L1 on MWV use. The speakers are from the following countries: China, Hong Kong and Macau (Chinese), Italy (Italian), and Argentina, Mexico, and Spain (Spanish).

In order to answer each of the research questions, different data from the sub-corpora described above were used. For RQ1, which describes MWV use in the TLC, data about speakers at the three levels of proficiency (i.e. B1, B2, and C1-C2) and from the three L1 backgrounds (i.e. Chinese, Italian and Spanish) were included. In addition, data from four tasks were used (i.e. presentation, discussion, interactive task, and conversation). It should be noted that data from the listening task was not included in this study given the limited amount of information that L2 speakers provide in such task (see section 3.5.3). For RQ2, which investigates the effect of learner variables on L2 speakers' MWV use, data coming from the three proficiency levels specified above, the three L1 backgrounds and two tasks (i.e. discussion and conversation) were used. Only the discussion and conversation tasks were taken into account to answer RQ2 as those two tasks occur across all proficiency levels. In regard to RQ3, which evaluates the effect of task type on MWV use, the data came from advanced C1-C2 speakers only given that the examination at the advanced level had the largest number and widest variety of speaking tasks (see section 3.5.3). Data from four tasks (i.e. presentation, discussion,

interactive task, and conversation) were included. As with the other research questions, speaker data from the three L1 backgrounds were also used to answer RQ3.

### 3.8. Limitations of corpus evidence in the TLC

As described in section 3.5.3, the tasks in the TLC are semi-formal in nature and the language sampled can be described as being "close to academic interaction" (Gablasova, Brezina & McEnery, forthcoming) due to the examination setting where the language was elicited. Because of these discourse characteristics, the corpus did not allow the observation of MWVs in more informal and less academic settings where they might occur more frequently (see section 2.3.1.3). A further limitation has to do with the overall low number of advanced speakers from the C2 proficiency band. Although combining C1 and C2 speakers was deemed a sensible decision to increase the size of the evidence at the advanced level, it was not possible to analyse differences in the MWV use between the most advanced (i.e. C1 and C2) speakers.

### 3.9. Accessing the corpus

The tool used to access the TLC was the Sketch Engine (Kilgarriff et al., 2014). The main reason for the selection of the Sketch Engine is that, at the time of this study, the TLC was only available via this tool. The Sketch Engine is a sophisticated corpus analysis tool originally launched in 2004. It allows users to create and manage their own corpora and to work with pre-loaded corpora. Although it was originally designed to help in lexicographic work, it has become widely used in a number of areas such as sociolinguistics, language teaching, discourse analysis, and translation (Kilgarriff et al., 2014). In order to analyse the TLC, the concordance tool, one of the Sketch Engine's

core functions, was used. The concordance tool allowed the visualisation of MWVs in context and provided the relevant metadata (i.e. information about the L2 speakers and the tasks where those MWVs occurred). The study also made use of other functionalities of the Sketch Engine which included Corpus Query Language (CQL) queries (see section 3.10.2), text type restricted searches (i.e. those based on particular variables of interest such as L2 proficiency, L1 background and task type), and frequency analyses showing the MWV frequency distribution per speaker.

## 3.10. Data collection

### 3.10.1. MWVs analysed in the study

The present study is based on the analysis of two types of MWVs: PVs and PPVs (see sections 2.2.2 and 2.2.3). First, a PV was understood as the combination of any verb proper or element with a verbal function and an adverbial particle from the ones listed by Quirk et al. (1985) and Huddleston & Pullum (2002) (e.g. *carry out*). These particles are listed in Table 3.5. Particle lists from those two works were selected as they are both comprehensive and have served as the reference particle lists in previous corpus-based studies on MWVs (e.g. Biber et al., 1999; Waibel, 2008; Gilquin, 2015b). Following Gilquin (2015b), one of the particles mentioned in Huddleston & Pullum's (2002) list (i.e. *home*) was excluded in this study on grounds of its lexical rather than grammatical status in current English.

Table 3.5 Adverbial particles in MWVs

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| aback | across | apart | astray | behind | forward(s) | out | through |
| about | after | around | asunder | by | in | over | to |
| aboard | ahead | ashore | away | down | off | past | under |
| above | along | aside | back | forth | on | round | up |

Second, a PPV was considered a combination of a verb proper or element with a verbal function, an adverbial particle from the ones listed in Table 3.5, and a preposition (e.g. *get along with*). Both transitive and intransitive PVs and PPVs are included in the analysis of data. Similarly, MWVs from all semantic categories were taken into consideration: fully, semi-, and non-compositional verbs.

Prepositional verbs (i.e. sequences of a verb + a preposition such as *look at*) were left out of the study. Verbs immediately followed by "unequivocal prepositions" (Quirk et al., 1985, p. 1151) such as *against, among, as, at, beside, for, from, into, like, of, onto, upon*, and *with* were considered prepositional verbs and thus excluded from the study. The same criterion for exclusion was applied to verbs followed by what Cowie & Mackin (1993) call compound prepositions, which include *ahead of, as far as, in front of, on top of,* and *out of* as well as those verbs followed by temporal and semi-temporal prepositions such as *during* and *before* (Lindstromberg, 2010). Despite being one type of MWV, prepositional verbs are beyond the scope of this study for two reasons. First, their overall frequency, semantics and syntax differ widely from those of PVs and PPVs (Lindstromberg, 2010; Biber et al., 1999) (see section 2.2.4) that they merit separate treatment. Second, from an SLA perspective, prepositional verbs are reported to pose different problems to L2 speakers than PVs and PPVs (Nesselhauf, 2005). While the difficulty with prepositional verbs tends to relate to form, namely, the selection of the

correct preposition, the main problem with PVs and PPVs has been claimed to lie at the semantic level and relates to their different levels of opaqueness (Waibel, 2008). All these properties make it worthwhile looking at prepositional verbs as a separate verb class.

### 3.10.2. Data extraction

The identification of MWVs in this study proceeded in four main steps: i) the extraction of PVs and PPVs using a tag-based query, ii) the expansion of the initial PV and PPV searches, iii) the estimation of the precision level of the corpus query, iv) and the manual scrutiny of results.

First, CQL-based tags[9] were used to automatically extract PVs and PPVs from the TLC. This method was favoured over an extraction of MWVs based on predefined lists of lexically specified MWV lemmas (e.g. Liu, 2011; Märzinger, 2013) or particles (e.g. González, 2010; Ke, 2013). A tag-based search guaranteed a high recall level of MWVs given that all possible forms tagged as verbs (V) followed by particles (RP) were retrieved. In contrast, a search based on previous lists of specific MWVs or particles would have resulted in a loss of data since only those verbs and particles in the list (and not any other possible PV or PPV form that the speakers produced) would have been retrieved. This would have rendered an incomplete account of L2 speakers' MWV use in the corpus and would certainly have had consequences on the frequency counts and estimation of the range of MWVs.

[9] The TLC was tagged using the CLAWS6 tagset (Garside, 1987).

In order to retrieve PVs and PPVs, the following case insensitive tag string was used: [tag= "(?i)V.*" & !tag="VM*"] []{0,3} [tag="RP*"] | [tag="(?i)V.*" & !tag="VM*"] []{0,2} [tag="RP*"] [tag="I.*"]. This tag search identified the verbal element (V.*) of PVs and PPVs followed by a particle (RP*) and a preposition (I.*) in the case of PPVs. It excluded modal auxiliaries and modal catenatives (VM*) (e.g. *can, will, ought*) given that those do not occur with particles (Waibel, 2008). Also, the tag search allowed for up to three intervening elements between the verb and the particle in PVs (e.g. *wake up, wake **her** up, wake **the children** up, wake **those three children** up*). This decision was deemed appropriate for three reasons. First, it helped to retrieve PVs with an adjacent or non-continuous particle. Second, PVs containing more than three intervening elements are reported to be rare (Gardner & Davies, 2007, 2018; Liu, 2011). Third, a systematic visual inspection of PVs in the TLC with four or more intervening elements revealed a high degree of error, i.e. the results did not correspond to PVs. Regarding PPVs, those with up to two intervening elements were searched for despite the fact that only a few of them take an object between the verb and the particle (see section 2.2.4). This was done to allow the retrieval of all possible PPV forms including those in which these intervening elements were repetitions, rephrases or a series of discourse markers, which were common given the spoken nature of the corpus data (e.g. *someone must have put **erm him** up to it*).

After the initial automatic extraction, the second step involved searching for sequences of verbs followed by prepositions or adverbs in order to maximise recall levels. The purpose of this expanded search was to trace cases of MWVs whose particles were mistagged as prepositions or adverbs as a result of POS-tagging errors. Since searching for all sequences of verbs + prepositions and verbs + adverbs in the TLC was

impractical and would have resulted in a large number of non-MWV hits, this extended search was based on Gardner & Davies (2007) core set of prepositions and adverbs reported to be likely to function as adverbial particles in MWVs. Using the grammatically tagged version of the BNC, Gardner & Davies conducted software queries to extract all lexical verbs followed by particles with up to two intervening elements (e.g. *take the nodules back*). As a result of these queries, they identified sixteen particles that appeared in the MWV forms found. These particles are shown in Table 3.6. As a follow-up step, the total number of tags that each of those sixteen forms received in the BNC was counted and later compared to the number of times that each form was tagged as a particle rather than as a preposition or another grammatical structure. The results of this comparison are also presented in Table 3.6.

Table 3.6 Gardner & Davies' (2007) list of adverbial particles in the BNC

| **Form** | **Total tags** | **# as adverbial particles** | **% as adverbial particles** |
|---|---|---|---|
| 1. out | 149,727 | 145,706 | 97.3 |
| 2. up | 180,792 | 158,064 | 87.4 |
| 3. down | 91,832 | 72,709 | 79.2 |
| 4. back | 97,154 | 75,223 | 77.4 |
| 5. off | 67,479 | 37,751 | 55.9 |
| 6. round | 30,821 | 10,895 | 35.3 |
| 7. along | 18,555 | 4,925 | 26.5 |
| 8. over | 128,304 | 32,526 | 25.4 |
| 9. around | 43,391 | 10,384 | 23.9 |
| 10. on | 705,790 | 54,956 | 7.8 |
| 11. through | 81,184 | 5,797 | 7.1 |
| 12. about | 190,615 | 12,587 | 6.6 |
| 13. in | 1,845,077 | 34,411 | 1.9 |
| 14. under | 60,049 | 313 | 0.5 |
| 15. by | 504,969 | 371 | 0.1 |
| 16. across | 24,053 | 13 | 0.1 |

Note. # = token frequency.

Out of the sixteen forms in Table 3.6, the extended search in the present study was based only on the first nine (i.e. *out, up, down, back, off, round, along, over,* and *around),* each of which was found to function as a particle more than 20% of the total number of times that they appeared in the BNC. The remaining seven forms were not considered for the extended search given the marginal number of times that they were reported to function as particles. Each of the nine selected particles was searched for individually by entering a tag string that helped to retrieve cases where those particles were tagged as prepositions or adverbs but not as adverbial particles (RP) since the latter were retrieved in the first search. For example, the particle *out* was searched for using the following tag string [word="out" & tag="I.*|R.*" & !tag="RP*"], where "I.*|R.*" refers to instances where *out* was tagged as a preposition or adverb and !tag="RP*" indicates that instances of *out* tagged as a particle should be excluded. The extended search retrieved a total of 971 occurrences. These were further coded manually as described in section 3.10.3.

The third step in the MWV identification process consisted of making sure that only PVs and PPVs made their way into the query results. To this end, the level of precision was calculated by dividing the number of correctly tagged MWV tokens, including both PV and PPVs, (i.e. 2369) by the total number of tokens tagged as MWVs (i.e. 7682). The resulting precision score was relatively low (31%). The reason for the low precision score could be explained by the fact that searching for an efficient method to automatic MWV extraction is a challenging task (Li et al., 2003; Villavicencio, 2003; Kim & Baldwin, 2010). On the one hand, particles tend to be tagged in various ways because they can have multiple functions depending on the context where they are used, and automatic taggers often fail to identify those functions. Thus, the same form (e.g. *on*)

can be tagged as a preposition when it should be tagged as a particle and vice versa (Waibel, 2008). On the other hand, the variability in the number and type of intervening elements in transitive PVs and PPVs presents significant problems to a fully automatic MWV extraction because MWVs cannot be retrieved, for instance, by using n-grams (Baldwin & Villavicencio, 2002; Baldwin, 2005). Therefore, as with almost any corpus query, the results of my search were not free from POS-tagging errors. In order to weed out false positives and improve precision levels, the results obtained after the first and expanded searches (steps one and two described above) were subject to manual scrutiny, which was the fourth and last step taken as part of the data extraction process. Details regarding the manual coding of the data as well as a description of the coding scheme are provided in the next section.

### 3.10.3. Manual coding of the data

For the manual analysis of the extracted data, a set of form-related, syntactic and semantic criteria was developed. The criteria consist of a series of guidelines that were used to analyse both PVs and PPVs. In some cases, certain criteria were not applicable to every MWV. For example, criterion h) below was used only in those cases where there were multiple overlapping concordance lines containing the same MWV form.

#### *3.10.3.1. Form-related criteria*

a) PVs are two-part verbs; they consist of **two** essential elements: a verb and a particle. These two elements might be contiguous or separated by one or more elements.
   Examples: *make out, pick it up, brought the children up*

Prepositional verbs, i.e. verb + preposition, were excluded from the data.

Examples: *The whole story is revolving around Greek gods.*

*I thought of you this morning.*

b) PPVs have **three** essential elements: a verb, a particle and a preposition.

Examples: *come up with an idea, get along with someone,* take your anger *out on* me

c) Nominal forms of PVs/PPVs including those functioning as objects of prepositions were included in the analysis.

Examples: […] *the government has banned cutting down trees* […]

[…] *can really help me mm er in finding out stuff* […]

d) Adjectival forms derived from PVs/PPVs were also included in the data.

Example: *gets messed out, a curled up cat*

e) All different variants of the same PV/PPV lemma were included in the data.

Examples: *goes up, going up, to go up, went up, gone up*

f) When two or more PV/PPV forms that differ at the lexical and/or grammatical level appeared next to each other, those forms were included in the data and counted separately.

Example: […] unemployment *w-was going up is growing up* and why […]

1 2

g) When two consecutive instances of the same PV/PPV form occurred without any interruptions, that PV/PPV was included in the data and counted only once.

Example: […] he *take up take up* the class […]

1

However, any intervening element between two instances of the same PV/PPV would cause the MWV form to be counted twice.

Example: […] *children often grew up* ***erm I*** *grew up but my dad* […]

1 2

h) Sometimes as a result of the automatic extraction of MWVs with intervening elements, the same MWV appeared in two or more overlapping concordance lines[10]. In these cases, the MWV was counted only once. Therefore, only one concordance line (usually the first one where the MWV appeared) was kept for the analysis.

1. *[…] but nowadays erm reform has been carried out mm and erm the payment […]*
2. *[…] but nowadays erm reform has been carried out mm and erm the payment […]*
3. *[…] but nowadays erm reform has been carried out mm and erm the payment […]*

[10] These were instances in which the MWVs were preceded by other verb forms as in *has been carried out*. In those cases, the Sketch Engine regarded the first element of the verbal structure (e.g. *has*) as the lexical verb of the MWV form, the last element as the particle (e.g. *out*), and any items interceding between those two as the intervening elements (e.g. *been carried*). While I acknowledge that the tag string used to identify MWVs could have been more specific to exclude cases where *have, be,* or *do* functioned as auxiliary verbs, I decided not to do it to be able to retrieve instances of MWVs where *have, be* and *do* did function as the lexical verbs of those MWV forms (e.g. *have around, be out for, do up*).

*3.10.3.2. Syntactic criteria*

i) Both transitive and intransitive PVs/PPVs as well as PVs/PPVs in passive voice constructions were included in the data set.

Example: […] *it is a very large and it <u>is made up of</u> glass* […]

j) To distinguish PVs (e.g. *run down a friend*) from superficially identical prepositional verbs (e.g. *run down the hill*), three syntactic tests were applied: particle fronting, particle movement, and question formation with fronted particles. Despite the variety of tests available to identify PVs (see section 2.2.2), the three above were selected following Biber et al. (1999) and O'Dowd (1998), who argue that those tests are sufficient to distinguish PVs from similar verb forms in most cases.

For intransitive PVs:

- *Particle fronting*

  In prepositional verbs, it is possible to place the preposition before the subject. This kind of fronting is not possible with particles in intransitive PVs.

  Examples: She climbed *up*. *<u>Up</u> she <u>climbed</u>.*

  *The tank blew up.* **<u>Up</u> the tank <u>blew</u>. (Quirk et al., 1985)**

  *He passed out.* **<u>Out</u> he <u>passed</u>. (Quirk et al., 1985)**

For transitive PVs:

- *Particle movement*

  The particle of a separable, transitive PV can usually appear either before or after the object noun phrase following the verb. On the contrary, prepositions can only precede the noun phrase.

Examples: *We turned off the light. We turned the light off. (O'Dowd, 1998)*

*We turned off the road. *We turned the road off. (O'Dowd, 1998)*

- *Question formation with fronted particles*

  Particles cannot be fronted and appear before interrogative words in wh-questions whereas prepositions can.

  Examples: *He called on the dean. On whom did he call? (Quirk et al., 1985)*

  *He turned the suspect in. *In whom did he turn?*

k) PVs/PPVs were not excluded from the data on the basis of any grammatical inaccuracies found in L2 speakers' production. For instance, all the following forms were considered valid PVs: *could came back, aim to came back, cames back*.

*3.10.3.3. Semantic criteria*

l) When neither of the previous syntactic tests helped to disambiguate the type of verbal form, i.e. to distinguish between a PV and a prepositional verb, the PV candidate was analysed in relation to its surrounding context. The PV status of a verb + particle combination was then checked using Cowie & Mackin's Oxford Dictionary of Phrasal Verbs (1993). If, according to its surrounding context, a PV was used with one of the meaning senses listed for such PV in the dictionary mentioned above, it was considered a PV. When there was not enough context to determine the meaning of the PV, that verb was excluded from the data. The Oxford Dictionary of Phrasal Verbs was selected not only because it draws on corpus evidence (i.e. the Oxford Corpus of the English Language) but also because it provides grammatical codes for each verb entry based on its possible meanings.

These grammatical codes specify whether a verb is followed by a particle, a particle and a preposition, or only a preposition depending on its intended meaning.

m) Semantic compositionality was not considered in the manual analysis. PVs and PPVs from all semantic categories (fully, semi and non-compositional structures) were included in the data. "Free combinations" (Quirk et al., 1985, p.1152), i.e. combinations where both the verb and the particle retain their distinctive meanings (e.g. *go out, fall down*), were also included in the analysis given that a clear-cut semantic differentiation between MWVs exhibiting different levels of transparency is usually unfeasible (see section 2.2.5) and beyond the scope of the present study.

n) Idiomatic expressions were left out in cases where the particle in them formed a unit with a following preposition and not with the verb.
Example: […] *he thought <u>out of the box</u>* […]

However, conventionalised idiomatic expressions where there was a separable, transitive PV in a syntactically fixed position were included in the data.
Examples: *I was crying my eyes out. (Quirk et al., 1985)*
*I was laughing my head off. (Quirk et al., 1985)*
*I was sobbing my heart out. (Quirk et al., 1985)*

### 3.10.4. Double coding of the data

Given that a few coding decisions during the manual analysis of the data were based on interpretive differences and relied on the rater's judgement, a sample of the data was double-coded in order to determine coding accuracy. Another coder's analysis of the

data proved to be both necessary and useful especially to confirm the MWV status of certain forms, which otherwise would be classified as other verb types if used in a different context and with a different meaning. This was particularly the case of polysemous verbs that may be classified as phrasal or prepositional depending on the meaning sense that they convey. For example, *'go through'* is regarded as a PV if it is used to express that something is completed or concluded successfully as in '*The proposal went through*'. However, the same verb form can be classified as a prepositional verb if used in a context where it means experiencing, enduring or suffering something (Cowie & Mackin, 1993) as in '*He had no idea of what they had gone through in life*'.

The double coding of the data was a three-step process that started with the selection of a random sample of the data using the sample function in the Sketch Engine. The sample contained 1500 MWVs, including both PVs and PPVs, which corresponded to 20% of the total number of MWVs that were extracted (i.e. 7682). The amount of data in the random sample was determined by the relatively high inference involved in the coding (Loewen & Plonsky, 2016; Brezina, 2018a). Once the data subset was obtained, the coding scheme, i.e. the criteria, was handed to the second rater (one with linguistic expertise) along with a print copy of Cowie & Mackin's (1993) dictionary. An explanation of the grammatical codes assigned to MWVs in the dictionary was also provided to the second coder. In the third step, the reliability of the coding was estimated by calculating the level of inter-rater agreement in the Agreement Calculator tool (Brezina, 2018b). Due to its reliability for categorical variables (Brezina, 2018a, 2018b), the inter-rater agreement statistic selected was Gwet's $AC_1$, which showed a high level of agreement between the raters ($AC_1 = 0.82$, $p < 0.001$). This level of

agreement was considered sufficient given the subjectivity involved in the coding. After a review of the discrepancies, no major patterns were found except the inconsistent application of syntactic criterion k) by the second rater (i.e. MWVs should not be excluded on the basis of any grammatical inaccuracies). These cases were later discussed and resolved by both raters.

### 3.11. Assessment of polysemy

In order to count the number of meaning senses expressed by polysemous MWVs, the following procedure was followed. First, from the MWVs that were included in this study after the manual analysis described in section 3.10.3, MWV lemmas that appeared at least twice were extracted (i.e. 168). MWVs that occurred once were excluded for polysemy assessment purposes since those were used to express one meaning only. Second, all instances where each MWV occurring at least twice appeared (i.e. 2223) were manually checked. The manual analysis consisted of examining the context where each MWV occurred and determining whether the meanings expressed were among the ones listed for such MWVs in the Oxford Dictionary of Phrasal Verbs (Cowie & Mackin, 1993). The reasons for selecting this dictionary are explained in section 3.10.3.3. Third, the meaning senses of a MWV that could not be interpreted or did not correspond to any of the ones listed for such MWV in the dictionary consulted were excluded from the analysis of polysemy.

### 3.12. Data analysis

For the quantitative analysis of the data, both descriptive and inferential statistics were used. The former allowed a description of general frequency patterns and helped to

understand how MWVs were distributed across L2 proficiency levels, L1 backgrounds and task types in the TLC. The specific descriptive statistics used in the study included i) raw frequencies of the most frequently occurring MWVs, lexical verbs and adverbial particles functioning in MWVs, and non-canonical MWV forms, ii) normalised frequencies of MWV occurrence, iii) measures of central tendencies (i.e. means), and iv) dispersion measures (i.e. standard deviations). The descriptive statistics were used to answer RQ1 and some parts of RQ2 and RQ3.

With respect to inferential statistics, two statistical tests were used in order to gauge the influence of different variables on MWV use. First, a two-way Analysis of Variance (ANOVA) was selected to answer RQ2. This test helped to determine the effect of individual learner variables as well as of their possible interaction on MWV use. In the test, L2 proficiency and L1 background were the independent variables whereas the frequency of MWV occurrence per thousand words was the dependent variable. This basis of normalisation (i.e. one thousand) was selected to best reflect the size of the samples coming from individual speakers. Second, a repeated-measures ANOVA was selected to answer RQ3. Given that RQ3 looked at the same advanced L2 speakers' performance in various tasks (i.e. multiple samples for each advanced L2 speaker were analysed), the repeated-measures ANOVA was considered the most suitable test (Brezina, 2018b). Because of violations of the assumption of sphericity, a Greenhouse-Geisser correction was used. For the purposes of carrying out the repeated-measures ANOVA, the type of speaking task was considered the independent variable and the frequency of MWV occurrence per thousand words the dependent variable. For both statistical tests, Bonferroni post-hoc tests were used.

It is important to point out that for the quantitative analyses, no distinction between PVs and PPVs was made. This decision was deemed appropriate due to the fact that looking at PPVs separately would not have rendered a sufficiently large amount of data for the quantitative analyses given the overall low occurrence of PPVs in the TLC in comparison to PVs. However, the overall normalised frequency of PPV occurrence and the raw frequencies of all PPVs identified are provided in section 4.2.1.

### 3.13. Summary

This chapter has provided a detailed description of the design and methodological aspects of this investigation. It first highlighted that the study is descriptive in nature and informed by both SLA and corpus-based research. Regarding its methodological approach, the study is based on L2-L2 comparisons. In other words, it compares MWV use as produced by different groups of L2 speakers. These groups differ from each other in terms of L2 proficiency and L1 background. Terminological aspects were also discussed in this chapter. Particular attention was given to the terms PV and non-canonical MWV. Despite the wide range of terms available, PV was selected to refer to combinations of verbs + particles since it is considered a theoretically neutral term and is commonly used in the literature about this verb type. With respect to non-canonical MWVs, reasons were provided as to why 'non-canonical' or 'non-standard MWV' is preferred over the term 'error' to refer to cases of unattested MWVs that the L2 speakers produced.

The chapter then proceeded to fully describe the spoken corpus on which this study is based, i.e. the Trinity Lancaster Corpus (TLC). In terms of speaker-related variables, the TLC contains information about L2 speakers from four different levels of

proficiency (i.e. B1, B2, C1, and C2) and from a variety of L1 backgrounds and age groups. Details regarding examiner-related variables were also provided and the format and discourse characteristics of the different speaking tasks were explained. The rationale behind the selection of the TLC relates to the multiple opportunities that it affords to analyse MWVs in spoken L2 communication and across a variety of L1 backgrounds and task types. Also, the explicit and reliable information about the speakers' level of L2 proficiency offered by the TLC was a key factor when selecting the corpus for this study. Despite the advantages of the TLC, some limitations of the corpus evidence were identified. These limitations include i) the semi-formal and close-to-academic discourse that the corpus contains which did not allow the analysis of MWVs in more informal spoken contexts and ii) the low number of C2 speakers. Following the introduction of the TLC, an overview of the sub-corpora used to answer the research questions was provided.

The last sections of the chapter described the corpus tool (i.e. the Sketch Engine) and the steps taken to identify MWVs in the TLC. Details regarding the manual analysis of data, the coding scheme, double-coding process, calculation of the inter-rater agreement level, and evaluation of polysemy were also provided. Finally, the chapter explained the different types of descriptive and inferential statistics used and outlined the ways in which they were computed.

# Chapter 4: Results

## 4.1. Introduction

This chapter presents the results obtained in the study introduced in Chapter 3. Findings are organised in terms of the three research questions that the study aimed to answer. First, section 4.2 outlines the results relating to overall MWV frequencies and usage patterns. It then goes on to report on the number of non-canonical MWV forms that L2 speakers produced and meaning senses associated with polysemous MWVs. Next, section 4.3 contains information regarding the effect of learner variables, i.e. L2 proficiency and L1 background, on the frequency and range of MWVs. Finally, section 4.4 presents the results regarding the effects of speaking task on L2 speakers' use of MWVs.

## 4.2. RQ1 L2 speakers' use of MWVs in the TLC

The aim of RQ1 was to provide a description of how MWVs, i.e. PVs and PPVs, are used by L2 learners in terms of i) MWV frequency and coverage, ii) lexical verb and particle productivity, iii) the number and type of non-canonical MWV forms, and iv) polysemy. The results pertaining to RQ1 are organised according to those main four aspects.

### 4.2.1. MWV frequency and coverage

The data analysis identified 2,349 MWV tokens and 294 distinct MWV lemmas of which 272 corresponded to PVs and 22 to PPVs. In terms of relative frequencies,

MWVs occurred 1,421 times per million words (pmw). PVs occurred 1,397 times and PPVs 24.2 times pmw. Regarding coverage, Table 4.1 shows that of the 294 MWV lemmas, a small set of 20 forms accounts for almost 60 per cent of all the MWV occurrences identified.

Table 4.1 Frequency and coverage of top 20 MWVs in the TLC

| MWV | Raw frequency | Percentage of all MWV occurrences in L2 sub-corpus |
|---|---|---|
| 1. go out | 280 | 11.91% |
| 2. grow up | 207 | 8.81% |
| 3. come back | 153 | 6.51% |
| 4. go back | 88 | 3.74% |
| 5. go on | 87 | 3.70% |
| 6. find out | 63 | 2.68% |
| 7. wake up | 61 | 2.59% |
| 8. come out | 55 | 2.34% |
| 9. get up | 54 | 2.29% |
| 10. give up | 40 | 1.70% |
| 11. take out | 36 | 1.53% |
| 12. move on | 32 | 1.36% |
| 13. pick up | 31 | 1.31% |
| 14. come on | 31 | 1.31% |
| 15. sum up | 29 | 1.23% |
| 16. keep on | 29 | 1.23% |
| 17. take off | 28 | 1.19% |
| 18. bring up | 27 | 1.14% |
| 19. make up | 26 | 1.10% |
| 20. go up | 26 | 1.10% |
| Total | 1,383 | 58.8% |

As can be seen in Table 4.1, the top three verbs, i.e. *go out, grow up,* and *come back,* are overwhelmingly more frequent compared to the other verbs within the top 20, whereas 126 verbs occurred only once (e.g. *flush out, look forward to, look out, get down*). The results also show that no PPVs made it to the top 20 verbs. Out of the 22 PPVs found, 14 occurred only once as shown in Table 4.2. The complete list of all the MWVs identified in this study along with their absolute frequencies is provided in the Appendix.

Table 4.2 PPV frequency

| PPV | Raw frequency |
|---|---|
| 1. come up with | 7 |
| 2. come out with | 5 |
| 3. go on with | 4 |
| 4. look out for | 2 |
| 5. go along with | 2 |
| 6. boil down to | 2 |
| 7. get back to | 2 |
| 8. look up to | 2 |
| 9. look forward to | 1 |
| 10. lead back to | 1 |
| 11. get up to | 1 |
| 12. get through to | 1 |
| 13. look down upon | 1 |
| 14. make up for | 1 |
| 15. clamp down on | 1 |
| 16. put up with | 1 |
| 17. go in for | 1 |
| 18. put down to | 1 |
| 19. bind up with | 1 |
| 20. rub off on | 1 |

| | |
|---|---|
| 21. hold on to | 1 |
| 22. keep up with | 1 |
| Total | 40 |

4.2.2. Lexical verb and particle productivity

The findings related to lexical verb and particle productivity can be divided into two categories: the frequency of lexical verbs and particles that participate in the formation of MWVs and the number of particles (in the case of lexical verbs) and verbs (in the case of particles) that they combine with. First, with respect to lexical verb frequencies, overall results indicate again that a small set of 10 lexical verbs is found in more than 60 percent of the MWV forms produced by the L2 speakers. In addition, more than half of the MWVs found in L2 production contain one of the five most frequent lexical verbs from this set (i.e. *go, come, grow, get*, and *take*). These findings are presented in Table 4.3.

Table 4.3 Most frequent lexical verbs functioning in MWV forms

| MWV | Frequency in MWVs | Percentage of all MWVs |
|---|---|---|
| 1. go | 534 | 22.73% |
| 2. come | 287 | 12.21% |
| 3. grow | 207 | 8.81% |
| 4. get | 112 | 4.76% |
| 5. take | 96 | 4.08% |
| 6. find | 63 | 2.68% |
| 7. wake | 61 | 2.59% |
| 8. give | 58 | 2.46% |
| 9. bring | 49 | 2.08% |
| 10. turn | 45 | 1.91% |
| Total | 1,512 | 64.31% |

Note: # = token frequency

In terms of particle frequency, Table 4.4 displays a frequency ranking of all the particles that occurred in the MWVs that the L2 speakers used. The table also shows the proportion (percentage) of adverbial particles out of all MWVs. It is clear from the values in the table that the particles *up* and *out* predominated in the MWVs given that more than half of the MWVs that the L2 speakers produced contained one of those two particles.

Table 4.4 Most frequent adverbial particles functioning in MWV forms

| | Adverbial particle | Frequency in MWVs | Percentage of all MWVs |
|---|---|---|---|
| 1. | up | 758 | 32.26% |
| 2. | out | 648 | 27.58% |
| 3. | back | 296 | 12.60% |
| 4. | on | 243 | 10.34% |
| 5. | down | 156 | 6.64% |
| 6. | off | 118 | 5.02% |
| 7. | in | 50 | 2.12% |
| 8. | over | 25 | 1.06% |
| 9. | around | 24 | 1.02% |
| 10. | along | 13 | 0.55% |
| 11. | by | 6 | 0.25% |
| 12. | away | 4 | 0.17% |
| 13. | through | 3 | 0.12% |
| 14. | about | 3 | 0.12% |
| 15. | round | 2 | 0.08% |
| 16. | forward | 1 | 0.04% |

Second, in order to capture the variation in the number and type of verb-particle combinations that the speakers produced, Table 4.5 provides information relative to the 10 most frequent lexical verbs functioning in MWV forms and the particles that they appeared with out of the 16 that occurred in the data. Findings show that the majority of the most frequent lexical verbs are also the most productive ones, i.e. they combined with the largest number of particles. For instance, *go* was not only the most frequent lexical verb but also the most productive one as it combined with 12 different particles. It is noteworthy, however, that there are three exceptions to this finding. The verbs *grow*, *find* and *wake*, which are among the 10 most frequent lexical verbs, occurred

with only one particle each to form three of the MWVs with the highest raw frequency (207, 63, and 61 respectively).

Table 4.5 Verb-particle combinations based on top 10 lexical verbs in MWV forms

| Lexical verb / Particle | go | come | grow | get | take | find | wake | give | bring | turn |
|---|---|---|---|---|---|---|---|---|---|---|
| up | 26 | 18 | 207 | 55 | 9 | - | 61 | 40 | 27 | 2 |
| out | 280 | 60 | - | 13 | 36 | 63 | - | 10 | 4 | 7 |
| back | 88 | 153 | - | 10 | 7 | - | - | 8 | 9 | 3 |
| on | 91 | 31 | - | 11 | 6 | - | - | - | - | 5 |
| down | 25 | 10 | - | 1 | 4 | - | - | - | 3 | 2 |
| off | 3 | - | - | 6 | 28 | - | - | - | - | 22 |
| in | 6 | 12 | - | 5 | 1 | - | - | - | 2 | - |
| around | 5 | - | - | 1 | - | - | - | - | - | 4 |
| over | - | 2 | - | - | 4 | - | - | - | 1 | - |
| along | 3 | 1 | - | 7 | - | - | - | - | - | - |
| by | 4 | - | - | 1 | - | - | - | - | - | - |
| away | - | - | - | - | 1 | - | - | - | - | - |
| through | 1 | - | - | 2 | - | - | - | - | - | - |
| about | - | - | - | - | - | - | - | - | 3 | - |
| round | 2 | - | - | - | - | - | - | - | - | - |
| forward | - | - | - | - | - | - | - | - | - | - |

Next, the prolific nature of particles was also examined to determine how many lexical verbs they combined with to create MWVs. The results of this analysis are presented in Table 4.6, which ranks particles according to the number of lexical verbs that they occurred with. As can be seen in the table, there are two major groups of particles. The first one containing highly productive particles (i.e. particles 1 to 9), particularly *up* and *out*. In contrast, the second group contains the particles whose productivity is rather low compared to the particles in the first group (i.e. 10 to 16).

Table 4.6 Particle productivity

| Adverbial particle | Number of lexical verbs these particles combined with to create MWV lemmas |
|---|---|
| 1. up | 68 |
| 2. out | 62 |
| 3. down | 31 |
| 4. off | 24 |
| 5. in | 19 |
| 6. on | 18 |
| 7. back | 15 |
| 8. over | 14 |
| 9. around | 11 |
| 10. along | 5 |
| 11. away | 3 |
| 12. by | 3 |
| 13. through | 2 |
| 14. about | 1 |
| 15. forward | 1 |
| 16. round | 1 |

Note: In order to calculate the number of lexical verbs that each particle occurred with, the same lexical verb that appeared in both a PV and PPV form was counted once (e.g. *come up* and *come up with*).

4.2.3. Number and type of non-canonical MWV forms

As described in the Methodology chapter (see section 3.3), in this study non-canonical forms are understood as lexical verb-particle combinations that are not attested in the sources (i.e. dictionaries) used for the manual analysis of the data. I acknowledge that from a more normative approach these verb forms might be labelled as errors or even be considered linguistic innovations (e.g. Schneider & Zipp, 2013 and Gilquin, 2015a). Nonetheless, because of the low frequency of the verb forms listed below, it was not possible to determine whether those forms occur on a systematic basis or show any

level of conventionalisation, which are two key criteria for identifying learner errors and linguistic innovations.

Table 4.7 shows the non-canonical MWV forms found in the data along with a description of the proficiency level and L1 background of the learners who used those forms. Out of the 20 non-canonical MWV lemmas identified, 18 occurred in the speech of a different L2 speaker. With the exception of *low down, change back, continue on,* and *lay around,* all forms occurred once. When two or more instances of the same verb were used, these were produced by the same speaker, with the verb *continue on* occurring in the speech of two different speakers from the same proficiency group (i.e. C1-C2) and L1 background (i.e. Spanish).

Table 4.7 Non-canonical MWVs

| | **Non-canonical MWV form** | **Raw frequency** | **L1 background** | **Proficiency level** |
|---|---|---|---|---|
| 1. | low down | 3 | Chinese | B2 |
| 2. | change back | 2 | Italian | B2 |
| 3. | continue on | 2 | Spanish | C2 |
| | | | Spanish | C1 |
| 4. | lay around | 2 | Spanish | B2 |
| 5. | become out | 1 | Italian | B2 |
| 6. | catch back | 1 | Italian | B2 |
| 7. | download down | 1 | Chinese | C1 |
| 8. | grow out[11] | 1 | Spanish | C1 |
| 9. | higher up | 1 | Chinese | C1 |
| 10. | lock off | 1 | Spanish | C1 |
| 11. | please up | 1 | Spanish | C2 |

[11] Notice that the prepositional verb *grow out of* is an attested form (Cowie & Mackin, 1993); however, *grow out* is not.

| | | | |
|---|---|---|---|
| 12. prefer out | 1 | Spanish | B1 |
| 13. remember back | 1 | Spanish | C2 |
| 14. restart over | 1 | Spanish | B2 |
| 15. return back | 1 | Chinese | B1 |
| 16. rush along | 1 | Chinese | B2 |
| 17. sack up | 1 | Italian | B1 |
| 18. say out | 1 | Chinese | C1 |
| 19. spend around | 1 | Spanish | B2 |
| 20. stay back | 1 | Spanish | B2 |

In terms of the verbs and particles in non-canonical MWVs, a relatively wide range of both was used. A comparison of the lexical verbs in non-canonical forms with the most frequent lexical verbs functioning in MWVs reveals that, with the exception of *grow*, none of the latter group was used in the non-canonical forms. Regarding particles, the table shows that *back* occurs with the highest frequency in non-canonical MWVs produced by speakers from all proficiency levels and L1 backgrounds. As far as L2 proficiency is concerned, speakers at the highest proficiency levels (B2 and C1-C2) produced more non-canonical forms (9 and 8, respectively) than the B1 speakers did (3). In terms of L1 backgrounds, 10 of the non-canonical forms come from the speech of Spanish speakers, followed by Chinese (6) and Italian (4) speakers.

A closer analysis of the data showed four recurrent patterns in the use of non-canonical MWVs. The first pattern relates to the use of an additional, redundant particle as in *return back, restart over, download down, continue on, higher up, low down,* and *remember back*. This pattern was found in the data from all proficiency levels. The following examples illustrate the redundancy of the particles given that the verbs that they accompany already express telicity or the meaning intended by the particles.

1. E[12]: Was it a good experience? So do you think you will do it again?
   C: yeah I hope I if I have time I will **return back**
   E: right if you have time… (Chinese, B1)

2. C: I get rid of I mean somebody into my house and took all all my stuff all my furniture everything so I have to **restart over** again… (Spanish, B2)

3. C: I've seen a lot of foreign websites and I really like them and about the songs that I wanted to
   E: mm
   C: **download it down** they… (Chinese, C1)

4. C: But maybe people always said we should **low down**[13] pollution… (Chinese, B2)

5. C: I will start my presentation by pointing out some negative effects of er piracy especially on the economic side er then I will **continue on** to speak about the organisation of this erm er activity… (Spanish, C2)

The second pattern refers to the use of forms in which the pronunciation of the lexical verb is close to one of other similar verbs in attested MWV forms. Thus, the lexical verbs, rather than the particles, seem to have been confused by L2 speakers. Two verbs illustrate this pattern: *sack up* and *lay around* as shown in Examples 6-7 below. After

[12] E: examiner, C: candidate
[13] Notice that the speaker appears to have used '*low*' as a verb rather than '*lower*'.

an analysis of the context and topic of the conversation where *sack up* in Example 6 was used, it seems that the intended MWV is likely *stack up,* which conveys the meaning of placing items in a pile. Similarly, *lie around* in Example 7, whose meaning is that of lying on a casual manner in a place (Cowie & Mackin, 1993), appears to be the intended MWV based on the context and topic of the interaction.

6. C: My mother and my brother erm **sack up** the shel= stack the shelves and I erm say er and I take er erm the cash out take cash out… (Italian, B1)

7. C: It'd be like wow I've found a million pesos what am I going to do with them? but first that kind of money doesn't just **lay around**… (Spanish, B2)

A third pattern was found in combinations where either a lexical verb was used with a particle close in meaning to the one intended (e.g. *lock off* instead of *lock out* in Example 8 and *grow out* instead of *grow up* in Example 9) or a particle was used as an aspectual marker to add a special meaning to the whole MWV combination. For instance, the analysis of the context surrounding the MWV in Example 10 shows that *up* was used to add a perfective meaning to the MWV. In Example 11, the particle *out* contributes the meaning of letting something in the open and making it public. In total, eight MWV forms were found that can be categorised as pattern three: *lock off, grow out, please up, say out, change back, prefer out, rush along,* and *spend around.*

8. C: The others er they just like er clo-closed the doors
   E: yes
   C: a-and **locked** them **off** … (Spanish, C1)

9. C: You should support her even though you don't like the music that that your little kid is listen but you don't I really

   E: sure

   C: that it's going only gonna be er for a while she's gonna **grow out** don't worry… (Spanish, C1)

10. C: It helps a people it helps somebody to to improve their abilities and try to try to erm try to win another person to **please** another person **up** … (Spanish, C2)

11. C: They are too shy to sh= and to act so even when they just just **say out** a word and even just they do action we should ou= show our praise … (Chinese, C1)

The fourth pattern was identified in forms where the L2 speakers combined lexical verbs and particles to express new meanings that could not have been expressed by using the lexical verbs alone. For instance, *become out* in Example 12, although not a plausible MWV, was used with the meaning of *popularise* (become popular). A similar pattern can be observed in Example 13 where another combination (i.e. *stay back*) is used as a synonym of *fall behind*. In addition to *become out* and *stay back*, one more non-canonical MWV illustrates the fourth pattern: *catch back*.

12. C: It's a sort of culture which must **become out** er a-and which must become er a subject th= erm in schools you know

E: mm

C: er so that people grow up with this erm with this feeling they have to do s= anything

E: mm

C: for the place they live… (Italian, B2)

13. C: They thought about the process of globalisation since the beginning because they didn't want to **stay er stay back** no they they want to continue to continue erm being a power in the in Asia in all the world and that's very important… (Spanish, B2)

4.2.4. Polysemy

Polysemy was assessed based on the number of meaning senses that were associated with each MWV lemma and that were attested in the dictionaries consulted in this study (see section 3.5.3.3). Only MWVs that appeared at least twice in the data were included for the purpose of evaluating polysemy. MWVs that occurred only once were excluded given that they were used to express one meaning only. The analysis showed that 168 MWV lemmas occurred at least twice, and those MWVs expressed 229 different meaning senses. This means that on average L2 learners are producing 1.36 meaning senses per one form of MWV. Table 4.8 displays the total number of meaning senses for the ten most polysemous MWVs as well as the number of times that each specific sense was used.

Table 4.8 Number of senses for top 10 polysemous MWVs

| MWV | Number of meaning senses | Frequency of individual meaning senses[14] |
|---|---|---|
| 1. come out | 8 | appear or emerge (22), become available to the public (15), be revealed or shown up clearly (4), be disclosed (2), disappear with cleaning (1), become visible from the sky (1), be spoken (1), become known or famous (1) |
| 2. come back | 7 | return to a place (141), return to previous state or action (4), become fashionable again (3), be restored (2), return to memory (1), return to a topic in a conversation (1) return successfully to prominence or fame (1) |
| 3. go on | 6 | continue an activity or relationship (34), take place, occur or happen (25), continue journey or career; to progress (15), continue by adding some new point to what has been said or written; to remark (7), continue speaking (3), pass by (2) |
| 4. go out | 6 | leave a place (e.g. one's home) for recreation or socialising (198), leave an area or situation (69), leave a place (more used as intensifier in the phrase *go out and do STH*[15]) (7), be broadcast, printed or distributed (2), end or close (2), stop burning or shining; being extinguished (1) |

[14] Notice that the sum of frequencies of individual meaning senses for a MWV sometimes does not match the total number of times that such MWV occurred (e.g. the sum of individual meaning sense frequencies for *come out* is 47, yet this MWV occurred 55 times). This is due to the fact that L2 speakers sometimes used the MWVs to express meanings that were not attested in the dictionaries consulted or that could not be interpreted and were thus excluded from the study.
[15] STH: something, SB: somebody

| | | |
|---|---|---|
| 5. pick up | 6 | take hold of and raise (14), take SB on board; stop to give a lift to SB (7), acquire knowledge of or a skill in STH, usually casually or without special study (4), make SB's acquaintance (2), collect STH (e.g. a parcel) from a place (1), acquire STH as one grows and develops (1) |
| 6. take off | 6 | remove clothing (esp. clothing) from one's body (8), cause an aircraft to leave the ground (7), remove an item from a menu in a restaurant (2), have some period as a break from work or holiday (1), lead, accompany SB away from one place to another (1), amputate (1) |
| 7. give up | 5 | leave or abandon STH (26), admit defeat or one's inability to do STH (5), hand over custody of SB (2), stop eating or drinking (2), sacrifice STH (2) |
| 8. make up | 5 | form or compose STH larger (11), invent STH (3), prepare a bed (2), apply cosmetics to one's face (2), make a decision (2) |
| 9. take up | 4 | adopt STH as a pastime (3), start a job; begin to work (2), adopt a role, position or attitude (1), move forward; advance in life (1) |
| 10. work out | 4 | exercise to improve health or strength (8), plan, devise or carefully think about STH (1), prove successful (1), find the answer to STH; solve (1) |

A comparison between the results in Tables 4.8 and 4.1 (frequency and coverage of top 20 MWVs) shows that eight of the most frequent MWVs are also among the most

polysemous ones, with the exception of *take up* and *work out*. Moreover, a closer examination of the verbs in Table 4.8 revealed that regardless of how polysemous the MWVs were, an average of one or two meaning senses were predominantly used. This means that only one or two senses accounted for 50% or more of the occurrences of each polysemous MWV form. As an example, the verb *go on* occurred 87 times, and it was associated with six different meanings. Despite its multiple senses, L2 speakers predominantly used *go on* to indicate (1) that an activity or relationship continues and (2) that an action or event takes place. Together both meaning senses made up 68% of the total occurrences of that verb in the data.

In order to further explore L2 speakers' semantic knowledge, the key senses (i.e. those accounting for at least 50% of the occurrences of a MWV) of the ten most polysemous MWVs were compared to the key senses reported for the same verbs in L1 speech. Information about key senses in L1 production comes from Liu & Myers (2018). Their study drew on data from the spoken and academic writing sub-corpora of the COCA corpus. Unlike previous studies reporting key meaning senses in L1 production (e.g. Garnier & Schmitt, 2015), Liu & Myers distinguish between key meaning senses of frequent MWVs in written and spoken registers. The meaning senses in L1 production used for comparison purposes in this section are those Liu & Myers report for spoken English. Table 4.9 presents the results of this comparison. Meaning senses are listed according to the percentage of coverage reported.

Table 4.9 Key senses of top 10 polysemous MWVs in L1 and L2 speech

| MWV | L1 speech | L2 speech |
|---|---|---|
| 1. come out | - Announce, display, or emerge from STH else (40.5%)<br>- Physically leave, exit, or erupt from a place or object (31.5%) | - Appear or emerge (47%)<br>- Appear in the shops; become available to the public (32%) |
| 2. come back | - Return to a place or a conversation topic (96.5%) | - Return to a place (92.1%) |
| 3. go on | - Happen, take place (76.2%) | - Continue an activity or relationship (40%)<br>- Take place, occur, happen (29%) |
| 4. go out | - Go on a date or to a specific location (42.5%)<br>- Take the field or go on a mission, often with a specific goal in mind (10.1%) | - Leave a place (e.g. one's home) for recreation or socialising (71%) |
| 5. pick up | - Get or take STH/SB from a place (44%)<br>- Learn, hear, or get wind of (19%) | - Take hold of and raise (48.3%)<br>- Take SB on board; stop to give a lift to SB (24.1%) |
| 6. take off | - Remove (34.7%)<br>- Leave the ground immediately (23.4%) | - Remove STH (esp. clothing) (40%)<br>- Cause an aircraft to leave the ground (35%) |

| | | |
|---|---|---|
| 7. give up | - Stop doing or having STH; abandon or surrender an activity, belief, possession (56.2%) | - Leave, abandon STH (e.g. job, career, activity) (70%) |
| 8. make up | - Create or invent (32%)<br>- Form a whole/compose (21%) | - Form, compose STH larger (55%) |
| 9. take up | - Discuss or deal with an issue, idea, matter, etc. (31%)<br>- Use a specific amount of time or effort (24.5%) | - Adopt STH as a pastime (43%)<br>- Start a job; begin work (29%) |
| 10. work out | - Happen or develop in a particular way, often successfully (45.5%)<br>- Solve a problem, plan, devise or think through STH (35.5%) | - Exercise to improve health or strength (73%) |

As can be seen in the table, with the exception of *take up* and *work out*, the majority of the key senses of the most polysemous MWVs that the L2 speakers produced tend to resemble those identified for the same verbs in L1 speech. In some cases, not only the type but also the number of key meaning senses reported for a particular MWV form in L2 production closely match those in L1 production (e.g. *come back*, *take off*, *give up*).

### 4.3. RQ2 The effect of learner variables

The second research question of this study sought to explore the effect of learner variables, i.e. L2 proficiency and L1 background, on L2 speakers' use of MWVs. As was reported in section 2.6, MWV use was operationalised in terms of two linguistic

variables: the frequency of MWV occurrence and the range of MWVs. The results of RQ2, therefore, will be organised according to those two variables and with respect to the level of L2 proficiency and L1 background of the speakers. Regarding the analysis of the data, a two-way analysis of variance (ANOVA) was conducted on the influence of L2 proficiency and L1 background (i.e. two independent variables) on the frequency of MWV use (i.e. dependent variable). L2 proficiency included three bands (B1, B2, C1-C2), and L1 background included three L1 groups (Chinese, Italian, and Spanish).

### 4.3.1. L2 proficiency

#### *4.3.1.1. Frequency of MWVs*

As far as the frequency of MWVs is concerned, the results indicate that MWVs occur relatively infrequently in L2 production. Despite this low frequency of occurrence, the results show a modest increase in the number of MWVs produced across proficiency levels, with MWVs appearing slightly more frequently in the speech of advanced (C1-C2) speakers. Table 4.10 provides the descriptive statistics showing general frequency patterns (normalised to 1,000 words) across the three proficiency levels.

Table 4.10 Frequency of MWV occurrence across proficiency levels

| Proficiency level | Number of speakers | Mean frequency | Std. Deviation |
|---|---|---|---|
| **B1** | 606 | 1.23 | 1.73 |
| **B2** | 502 | 1.26 | 1.49 |
| **C1-C2** | 240 | 1.58 | 1.65 |

The two-way ANOVA revealed that there was not a statistically significant interaction between the effects of L2 proficiency and L1 background [$F$ (8,1339) = 0.681, $p > 0.05$,

partial eta$^2$ = 0.002]. Similarly, the main effect for proficiency on the frequency of MWVs was also found to be statistically not significant [*F* (2, 1339) = 2.803, p > 0.05, partial eta$^2$ = 0.004]. In contrast, a statistically significant effect for L1 background on the frequency of MWVs [*F* (2, 1339) = 14.542, p < 0.001, partial eta$^2$ = 0.021] was found. The small effect size indicates that only 2.1% of the variation in the number of MWVs is explained by the influence of L1 background. Post hoc Bonferroni comparisons showed that there was a statistically significant difference between Chinese and Italian speakers (p < 0.001) and between Chinese and Spanish speakers (p < 0.001).

*4.3.1.2. Range of MWVs*

Despite the fact that L2 proficiency was not found to be a statistically significant factor in MWV use, a more detailed look at the verbs produced by B1, B2 and C1-C2 speakers revealed different patterns of MWV use across proficiency bands. With respect to the range of MWVs, the findings can be divided into three categories: a) the most frequent MWVs produced by L2 speakers across the three proficiency levels, b) the range of the most common lexical verb lemmas functioning in MWV verbs, and c) the specific set of the most frequent MWVs per proficiency band. First, in terms of the most frequent MWVs across proficiency bands, Figure 4.1 shows the top 25 MWV lemmas and their distribution per proficiency level out of a total of 238 lemmas found in the conversation and discussion tasks.

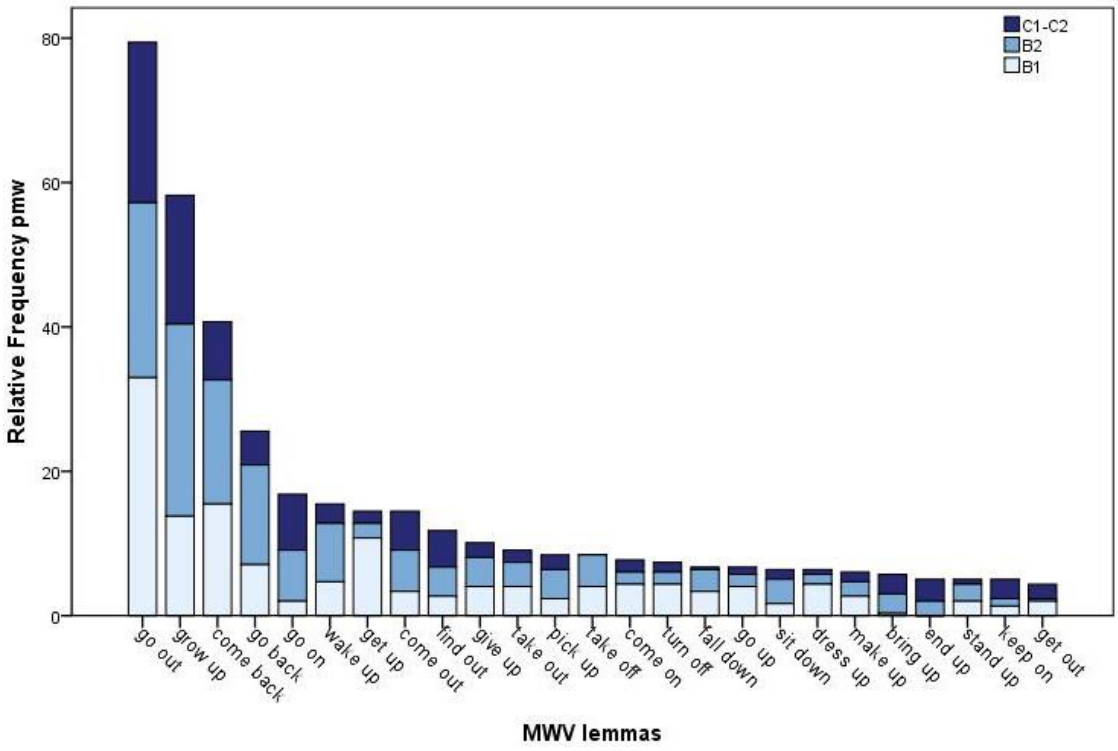


Figure 4.1: Top 25 MWVs in discussion and conversation tasks

What the figure shows is a near Zipfian distribution of the top 25 MWVs, with a small number of high-frequency verbs making up a large proportion of the MWV occurrences in L2 speech. Moreover, the figure shows that there is a high degree of inter-group variation. All 25 MWVs are unevenly distributed among the three proficiency groups, with specific MWV forms occurring very frequently in the speech of only one group of learners, e.g. *go out*, *get up*, *turn off*, and *dress up* in B1 production, and *grow up*, *go back*, and *wake up* in B2 speech. Still other verbs were absent in the L2 production of a particular group. For instance, *end up* was found in B2 and C1-C2 speech, yet it did not occur in beginner L2 production, whereas *take off* was only produced by beginner (B1) and intermediate (B2) speakers.

As part of the analysis of data, inter-speaker variation was considered in greater detail. Findings in this respect revealed a relatively high degree of variation within the same

proficiency group. For example, while 304 speakers from the B1 band did not produce any MWV in their interactions, others produced up to 15 MWVs per 1000 words. The same holds true for B2 and C1-C2 bands. B2 speakers used from 0 to 9 MWVs, and C1-C2 speakers from 0 to 8 MWVs per 1000 words.

A closer look at the most prolific speakers in each proficiency band, i.e. those producing 8 or more MWVs per 1000 words in conversation and discussion tasks, showed that high MWV frequency was not always equivalent to a wide variety of MWV lemmas. In fact, in most cases, fifty percent or fewer of the MWVs produced by those prolific speakers corresponded to distinct MWV lemmas. The analysis of data even revealed cases where all instances of the MWVs that a prolific speaker produced corresponded to only one MWV lemma.

Second, findings regarding the lexical verbs used in MWVs in each proficiency level indicate that overall the range of lexical verbs increases with higher proficiency. While B1 and B2 speakers used 66 and 75 different lexical verbs respectively, advanced C1-C2 speakers exhibited the most varied range with 83 different lexical verbs. Upon a closer inspection of the most common lexical verbs functioning in MWVs, however, the results showed that a small group of five verbs was responsible for the formation of a large number of MWVs in B1, B2 and C1-C2 speech and that this set of verbs is similar across proficiency bands. More than 50 percent of all the MWVs that B1, B2 and C1-C2 learners produced in the discussion and conversation tasks contained one lexical verb from this small group. What is more, *go* and *come* are ranked as the first and second most frequent lexical verbs forming MWVs in all proficiency levels. These results are visualised in Table 4.11. A more detailed look at the verbs in the table also

revealed that *go*, *come*, *get*, and *take* joined the largest number of particles to create MWVs, *grow* being the exception as it joined only one particle (i.e. *up*) across all levels.

Table 4.11 Most frequent lexical verbs in MWVs per proficiency level

| | **Go** | | **Come** | | **Grow** | | **Get** | | **Take** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Freq.** | **%** | **Freq.** | **%** | **Freq.** | **%** | **Freq.** | **%** | **Freq.** | **%** |
| **B1** | 143 | 24.4% | 78 | 13.3% | 41 | 7.0% | 58 | 9.9% | 35 | 5.9% |
| **B2** | 155 | 24.5% | 80 | 12.6% | 79 | 12.5% | 18 | 2.8% | 30 | 4.7% |
| **C1-C2** | 119 | 25.4% | 55 | 11.7% | 53 | 11.3% | 17 | 3.6% | 12 | 2.5% |

Note: Freq. = frequency of lexical verb in MWVs; total MWV occurrences in B1 (585), B2 (632), C1-C2 (468).

Third, overall findings related to the range of MWVs used by the L2 speakers show that MWV lemmas also become more varied as L2 proficiency increases. C1-C2 speakers produced a total of 138 distinct MWV lemmas followed by B2 and B1 speakers who used 132 and 122 MWV lemmas respectively. Despite this variety in MWV lemmas, a finer-grained examination of the 25 most frequent MWVs per proficiency band showed a high degree of overlap among the three proficiency groups. In other words, the set of the most frequent MWVs remains largely unchanged in B1, B2, and C1-C2 speech. Fourteen out of the 25 most frequent MWV lemmas that B1 learners produced also occurred very frequently in B2 and C1-C2 speech. In Table 4.12, those 14 MWVs are italicised, and the MWVs that occurred with a high frequency in the speech of a particular group of speakers are highlighted in bold.

Table 4.12 Top 25 MWVs per proficiency level

| B1 | B2 | C1-C2 |
|---|---|---|
| *1. go out* | *1. grow up* | *1. go out* |
| *2. come back* | *2. go out* | *2. grow up* |
| *3. grow up* | *3. come back* | *3. come back* |
| *4. get up* | *4. go back* | *4. go on* |
| 5. go back | *5. wake up* | *5. come out* |
| *6. wake up* | *6. go on* | *6. find out* |
| **7. dress up** | *7. come out* | *7. go back* |
| *8. come on* | 8. take off | 8. end up |
| 9. turn off | *9. pick up* | **9. keep on** |
| *10. give up* | *10. find out* | 10. bring up |
| *11. take out* | *11. give up* | *11. wake up* |
| 12. take off | *12. sit down* | **12. show off** |
| **13. go up** | *13. take out* | 13. get out |
| *14. come out* | 14. fall down | *14. give up* |
| 15. fall down | 15. bring up | *15. pick up* |
| *16. find out* | 16. stand up | *16. come on* |
| 17. make up | *17. get up* | *17. get up* |
| *18. pick up* | 18. make up | **18. give out** |
| **19. break up** | 19. go down | *19. take out* |
| *20. go on* | 20. end up | 20. go down |
| 21. stand up | 21. carry on | **21. put down** |
| 22. get out | *22. come on* | *22. sit down* |
| 23. carry on | **23. throw up** | **23. come in** |
| **24. get on** | 24. turn off | **24. hang out** |
| *25. sit down* | **25. figure out** | **25. look up** |

It should be noted that six MWVs (i.e. *come out, come back, go on, go out, pick up,* and *give up*) of the 14 that occurred very frequently in all proficiency levels are the most polysemous of all the MWVs found in the corpus. The number of meaning senses per each of those six MWVs ranges from five to eight (see section 4.2.4). The 14 shared

MWVs were explored further to determine if the variety of meaning senses increases as speakers become more proficient. Results in this respect showed that the number of meaning senses per MWV increased slightly with higher proficiency in six out of the 14 shared verbs: *go on, pick up, find out, come out, wake up,* and *come back*. The difference being particularly prominent between the beginner B1 level and the other two proficiency groups (B2 and C1-C2). In the other eight cases, the number of meaning senses remained the same across proficiency levels.

An analysis of the MWVs that occurred with high frequency in the speech of only one proficiency group provided some evidence that MWV frequency is affected by the topic of the interactions in which speakers participate. This is particularly the case at B1 and B2 levels. Five of the six MWVs in bold in the B1 and B2 columns were used by speakers in connection with the specific topics that were raised in their respective examinations. The following examples illustrate how a specific topic appears to have influenced the high frequency of MWVs in B1 and B2 speech. The topics discussed in the examples are fashion (14-15), relationships (16-17), and pollution and recycling (18-20)[16].

14. C: I am going to talk about cosplay cosplay

    E: yeah

    C: have you ever heard about?

    E: No I haven't no no

[16] In examples 18 to 20, although they are all clearly related to the topic of pollution, speakers used the particle '*up*' when the context suggests a different particle would have been more appropriate (e.g. *out, away*).

C: Okay er well basically like people that **dress up** like er different characters like a Halloween costume… (Spanish, B1)

15. C: I like the clothes er of this time
    E: mm
    C: but my mum say me that when she was a little girl she was erm **dress up** with the same clothes… (Spanish, B1)

16. C: For example er two years ago erm my dad and his his girlfriend **broke up** so and my dad was so sad… (Spanish, B1)

17. C: I met my girlfriend in er in er London
    E: mm
    C: and erm we ber= we **break up** just yesterday so… (Italian, B1)

18. C: People doesn't care about it they always like in the car they **throw up** trashes sorry **throw up** trash and and… (Spanish, B2)

19. C: They **throw up** something in the street or anything… (Spanish, B2)

20. C: I will charge for example if someone **throw up** like erm bottles in the s-street or in the way… (Spanish, B2)

The analysis of the MWVs that occurred with high frequency in advanced C1-C2 speech only, nonetheless, indicates that their use is not necessarily associated with a

specific topic. The following examples illustrate the use of *show off* in three different conversations between candidates and examiners. Information about the topics for each conversation is provided below each example.

21. C: It's not a religious r festival any more it's just become a festival of er **showing off** with horses… (Italian, C2)

    *Topic: Festivals and celebrations in Italy*

22. C: So it's status they always **show off** if they have a new designer bag even though it's not original… (Spanish, C1)

    *Topic: Designer goods*

23. C: In the USA that's that happens or in very developed countries that usually happens

    E: yeah

    C: That people want to **show off** and show and and they'll **show off** with the bandage

    E: yes

    C: on their nose like

    E: mm

    C: I paid for this… (Spanish, C1)

    *Topic: Plastic surgery*

### 4.3.2. L1 background

#### *4.3.2.1. Frequency of MWVs*

For the study of L1 background and its effect on L2 MWV use, three L1 groups (Chinese, Italian, and Spanish) were analysed. As mentioned in the Methodology chapter (see section 3.7), those three L1 groups were selected given that they offer the best L1 representation in the corpus and account for 70% of all speakers in the TLC.

The results indicate that both the frequency of MWVs and frequency variation across L1 backgrounds is rather low as shown in Table 4.13. The descriptive statistics providing information about general frequency patterns across L1 backgrounds are presented in Table 4.13. Results are again normalised to 1,000 words to best reflect the size of the samples.

Table 4.13 Frequency of MWV occurrence across L1 backgrounds

| **L1 background** | **Number of speakers** | | | **Mean frequency** | **Std. Deviation** |
|---|---|---|---|---|---|
| | **B1** | **B2** | **C1-C2** | | |
| **Chinese** | 108 | 96 | 33 | 1.92 | 2.06 |
| **Italian** | 149 | 129 | 68 | 1.22 | 1.51 |
| **Spanish** | 349 | 277 | 139 | 1.15 | 1.49 |

#### *4.3.2.2. Range of MWVs*

Findings related to the range of MWVs per L1 background are reported based on B1 speakers' use of MWVs. The B1 group was selected as it contains the largest number of L2 speakers in the sub-corpus used in this study (see section 3.7). Range was

operationalised in terms of i) the core set of the most frequent MWVs across L1s, ii) the variety of lexical verbs used in MWVs per L1 background, and iii) the most frequent MWVs produced by each L1 group. First, with respect to the most frequent MWVs across L1 backgrounds, Figure 4.2 displays the top 25 MWVs and their distribution across L1s. As can be seen in the figure, together the top 25 MWVs account for nearly 70% of the MWV occurrences in each L1 group. Although B1 speakers produced a relatively low number of MWVs (585), inter-group variation can still be observed in the figure. The majority of verbs that made it to the top 25 are unevenly distributed, with MWVs occurring mostly in the speech of Spanish speakers, who not only produced the largest number of MWV tokens (289) but also of MWV lemmas (82). Due to the overall low frequency of MWVs in B1 data, however, differences in the distribution of MWVs per L1 background are rather small, particularly at the tail of the distribution as the figure shows.

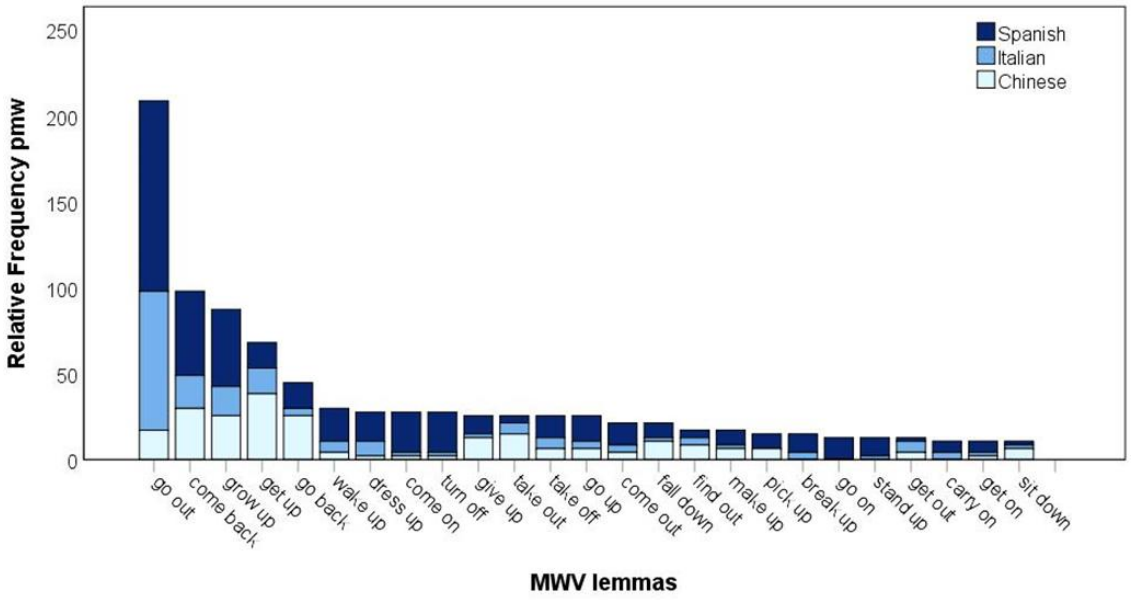


Figure 4.2: Distribution of top 25 MWVs per L1 background

Inter-group variation was also reflected in the number of speakers from each L1 background whose production did not include any MWVs. While 32.4% of Chinese B1 speakers did not use MWVs in their interactions, more than half of Italian speakers (53%) and of Spanish speakers (54.4%) did not produce any MWV at all.

Second, the study looked in detail at the range of lexical verbs that each L1 group used to create MWVs. Spanish speakers exhibited the widest variety of lexical verbs (49) followed by Italian and Chinese speakers who used 33 and 31 different lexical verbs respectively. As far as the most frequent lexical verbs are concerned, the results line up fairly equally with the findings based on L2 proficiency. A small set of five lexical verbs is found in the majority of MWVs that each L1 group produced as shown in Table 4.14. Not only are those lexical verbs the most frequent, but they are also the most productive in all L1 groups. With the exception of *grow*, all verbs combined with at least three and up to nine different particles.

Table 4.14 Most frequent lexical verbs in MWVs per L1 background

| | **Go** | | **Come** | | **Grow** | | **Get** | | **Take** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Freq.** | **%** | **Freq.** | **%** | **Freq.** | **%** | **Freq.** | **%** | **Freq.** | **%** |
| **Chinese** | 23 | 13.9% | 23 | 13.9% | 12 | 7.2% | 27 | 16.3% | 16 | 9.6% |
| **Italian** | 45 | 34.3% | 13 | 9.9% | 8 | 6.1% | 12 | 9.1% | 7 | 5.3% |
| **Spanish** | 75 | 25.9% | 42 | 14.5% | 21 | 7.2% | 19 | 6.5% | 12 | 4.1% |

Note: Freq. = frequency of lexical verb in MWVs; total MWV occurrences in B1 Chinese (165), Italian (131), Spanish (289).

Next, the MWVs produced by each L1 group were examined. Overall findings in this respect indicate that Chinese and Italian speakers performed similarly in terms of the

number of MWV lemmas that they used (53 and 51 respectively). However, Spanish speech contained the largest number of distinct MWV lemmas (82). A further analysis of the top 25 most frequent MWVs in each L1 group showed a much lower degree of overlap compared to the one observed in MWV use across proficiency levels. Speakers from all three L1s used a common set of only eight MWVs very frequently (i.e. *get up, come back, grow up, go back, go out, fall down, take off,* and *go up*) as can be seen in Table 4.15, whereas eleven MWVs occurred in two out of the three L1 groups. In Table 4.15, the eight MWVs that were found in the production of Chinese, Italian and Spanish B1 speakers are italicised, and MWVs in bold correspond to those that occurred with high frequency in the speech of one particular L1 group.

Table 4.15 Top 25 MWVs per L1 background

| **Chinese** | **Italian** | **Spanish** |
|---|---|---|
| *1. get up* | *1. go out* | *1. go out* |
| *2. come back* | *2. come back* | *2. come back* |
| *3. grow up* | *3. grow up* | 3. *grow up* |
| *4. go back* | *4. get up* | **4. come on** |
| 5. *go out* | 5. dress up | **5. turn off** |
| 6. take out | 6. wake up | 6. wake up |
| 7. give up | *7. take off* | 7. dress up |
| *8. fall down* | **8. look up** | *8. go up* |
| **9. set off** | 9. get out | *9. go back* |
| 10. find out | 10. take out | *10. get up* |
| **11. take down** | 11. come out | 11. come out |
| 12. pick up | 12. find out | *12. take off* |
| **13. set up** | 13. carry on | **13. go on** |
| **14. come in** | 14. break up | **14. send off** |
| **15. get in** | **15. stay on** | **15. stand up** |
| **16. sit down** | *16. go up* | 16. break up |

| | | |
|---|---|---|
| 17. make up | **17. do up** | 17. give up |
| *18. take off* | *18. go back* | 18. pick up |
| **19. write down** | **19. keep on** | *19. fall down* |
| **20. calm down** | **20. pay back** | 20. make up |
| *21. go up* | **21. take back** | **21. get on** |
| **22. stay out** | **22. sort out** | **22. throw up** |
| **23. stay up** | **23. go in for** | **23. work out** |
| **24. come out with** | **24. try out** | 24. carry on |
| 25. get out | *25. fall down* | **25. take on** |

Before analysing the MWVs that occurred exclusively in the speech of a particular group of speakers, let us briefly recap that B1 learners participate in two tasks: discussion and conversation. In the first of these tasks, speakers discuss a topic of their choice that they previously prepared. In the conversation phase, they talk about two subject areas selected by the examiner. Upon closer inspection, the MWVs in bold (i.e. those that occurred with high frequency in specific L1 groups) are often used in connection with the variety of topics that speakers chose for the discussion phase of their examination. Therefore, the MWVs vary from one L1 group to the other.

The MWVs in bold used in the discussion task, however, can be grouped into two types of MWVs. The first type of verbs includes those used repeatedly by the same speaker and in association with a particular topic. Such is the case of *set off [fireworks]* used to discuss the topic of festivals and celebrations in China, *take down [an enemy]* within the context of fighting video games, and *do up [one's laces]* to talk about hip-hop fashion. These cases are illustrated in Examples 24-26.

24. C: This year we we didn't **s-set off** any fireworks because I think if we **set off** a lot of fireworks our area will be very dirty […] I'm very happy just watching the fireworks another people have **set off** … (Chinese, B1)

25. C: You need to win the round and win the battle er your erm erm your direction is to kill the enemy or or **take down** the enemy or or **take down** the bomb or security's room […] and don't like attack in the room and to take out the take out the hostage or **take down** the bomb… (Chinese, B1)

26. C: You should wear your trainers without **doing** your laces **up** […] usually the erm the erm trainers are erm erm erm

    E: mm

    C: erm because erm it's erm a style of er hip-hop er er wear your trainers without **doing** your laces **up**… (Italian, B1)

The second type of verbs is those whose use does not appear to be related to a specific topic given that they were used by different speakers from the same L1 group to talk about a variety of topics in the discussion phase. Rather than a single topic being consistently associated with their use, it is the fact that these MWVs can be used across topics and contexts what seems to influence their high frequency in each L1 group. Examples of this type of MWVs and the topics of the discussions in which they were used include *calm down* (topics: how to pass an English test, instrumental music), *come on* (topics: playing video games with friends, horror films, going to Guadalajara), *send off* (topics: playing games, FIFA rules), and *get on* (topics: working as a salesman, Carly and Friends TV show).

Among the high frequency MWVs that occurred in the speech of only one L1 group is another set of MWVs whose use appears to be related to the topics raised by examiners in the conversation phase of the examinations. In each L1 group, there is a set of MWVs that recurred when speakers talked about certain subject areas selected by examiners in the conversation task. Even though these MWVs are different for each L1 group, they are all part of a semantically related set as the examples in Table 4.16 illustrate.

Table 4.16 High-frequency MWVs in the conversation task

| **Topics** | **MWVs** | **Examples** |
|---|---|---|
| Rules and regulations at school/ home | come in<br>sit down<br>stay out<br>stay up<br>turn off<br>stand up | C: Sometimes we will er teacher maybe will let you **come in** but after class er she will tell er tell y= tell y= tell you… (Chinese, B1)<br><br>C: Ah what's the rules? erm every everyone know children shouldn't shouldn't **stay out stay out**… (Chinese, B1)<br><br>C: They student er have to er **turn off** their mobile phone erm… (Spanish, B1)<br><br>C: We can't talk<br>E: mm<br>C: without permission and we **can't stand up** without permission… (Spanish, B1) |
| Learning a foreign language | write down<br>come out with<br>look up<br>sort out | C: In our school our teacher always erm teach us English and do some exercises<br>E: mm<br>C: always let us to erm **write down** some new words every day… (Chinese, B1) |

| | | |
|---|---|---|
| | | C: If you don't know a word you ma= you should er **look** it **up** a dictionary… (Italian, B1)<br><br>C: The best way is er is ah er to study the grammar and er to and to **sort out** to to speak… (Italian, B1) |
| Health and fitness | stand up<br>work out | C: You have to eat er healthy and if you no if you practise any sport or<br>E: mm<br>C: if you you can't **work out**… (Spanish, B1)<br><br>C: It doesn't matter if you are in a very er small space you have to move wake up er **stand up** go around… (Spanish, B1) |

### 4.4. RQ3: The effect of text-specific variables

RQ3 investigated the effect of text-specific variables, i.e. type of speaking task, on L2 speakers' use of MWVs. As in the previous section, the results regarding RQ3 are reported in terms of the frequency of MWV occurrence and the range of MWVs found in the production of advanced L2 speakers given that they participate in the largest number of speaking tasks (see section 3.5.3).

#### 4.4.1. Frequency of MWVs

In order to answer RQ3, advanced L2 speakers' (n=240) use of MWVs was evaluated across four tasks: presentation, discussion, interactive task, and conversation. Table

4.17 displays the descriptive statistics for the frequency of MWVs across all four tasks. Normalised MWV frequencies per speaker were computed per 1000 words to best reflect the size of the samples.

Table 4.17 Frequency of MWVs across tasks

| Task | Mean frequency | Std. Deviation |
|---|---|---|
| **Presentation** | 2.45 | 2.94 |
| **Discussion** | 1.57 | 2.29 |
| **Interactive task** | 1.34 | 2.64 |
| **Conversation** | 1.60 | 2.10 |

As the table shows, the monologic presentation is the task where speakers produced more MWVs per 1000 words, followed by conversation and discussion, with the interactive phase being the task where L2 speakers produced the lowest number of MWVs. A repeated-measures ANOVA with a Greenhouse-Geisser correction for sphericity determined that mean MWV frequency differed statistically significantly between tasks [$F$ (2.743, 655.656) = 11.023, $p < 0.001$, partial eta$^2$ = 0.044]. Post hoc tests using the Bonferroni correction showed a statistically significant difference between the presentation and every other task (all $p < 0.001$). There were no differences between any of the remaining three tasks (all $p = 1.0$).

Table 4.18 provides an overview of the distribution of MWVs across the four tasks as well as their coverage in C1-C2 production. Overall, the speech of advanced C1-C2 speakers contained 989 MWV tokens of which almost 40% were elicited in the presentation task alone. Advanced L2 production also contained the largest number of

PPV occurrences (e.g. *come up with*), that is 28 out of a total of 40. The presentation task contained not only most of those 28 PPV occurrences (11) but also the largest number of PPV lemmas (8).

Table 4.18 Distribution and coverage of MWVs across tasks

| **Task** | **Number of MWV tokens** | **Percentage of all MWVs in C1-C2 speech** |
|---|---|---|
| **Presentation** | 386 | 39.0% |
| **Discussion** | 192 | 19.4% |
| **Interactive task** | 135 | 13.6% |
| **Conversation** | 276 | 27.9% |
| *Total* | *989* | *100%* |

4.4.2. Range of MWVs

The range of MWVs was examined with respect to the following two features: i) the variety and distribution of the MWVs with the highest frequency across tasks and ii) the specific set of the most frequent MWVs per task. In order to explore the first aspect, the top 25 MWVs across tasks were analysed as Figure 4.3 shows. Frequencies are normalised to 100,000 words to reflect the size of the sub-corpora.

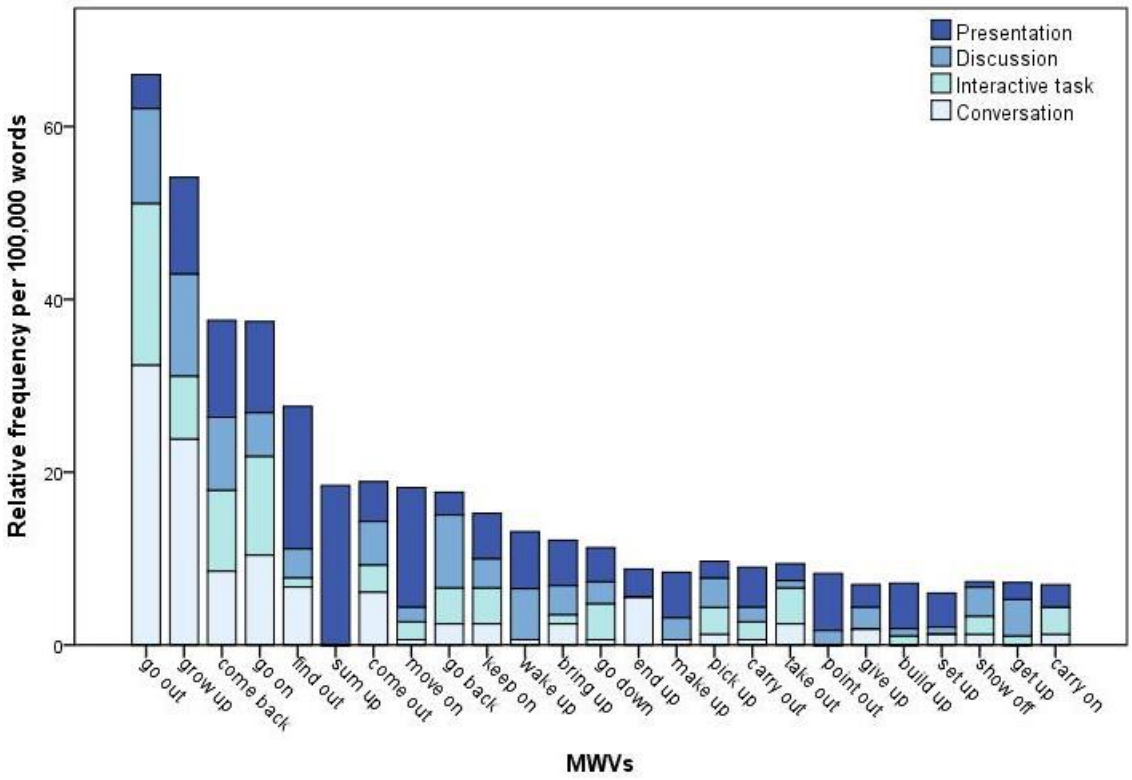


Figure 4.3: Top 25 MWVs in C1-C2 speech across tasks

Two key findings stand out from Figure 4.3. First, together the MWVs in Figure 4.3 account for 61.6% of all MWV occurrences in C1-C2 speech. In other words, a small group of 25 MWVs is responsible for most MWV occurrences in the data across all four tasks. Second, despite their overall high frequency in advanced C1-C2 production, those top MWVs are very unevenly distributed, with some verbs being predominantly used in the conversation section (e.g. *go out, grow up, come out, end up*), others in the discussion (e.g. *go back, show off, get up*), and still others in the interactive task (e.g. *go down, take out*). It should be noted that this unequal distribution is even more evident between the presentation and every other task. All 25 MWVs occurred at least once in the monologic presentation while some of them were not used at all in one or more of the other three dialogic tasks (e.g. *wake up, point out, build up*). Moreover, 13 out of

the top 25 MWVs occurred more frequently, and even exclusively, in the presentation task (e.g. *sum up, find out, move on, make up, set up*).

As regards the variety of MWVs used in each task, results indicate that the presentation elicited the widest range of MWV lemmas (127) followed by the conversation (91) and the discussion (88). As expected, the interactive phase, whose overall MWV count is lower compared to the other tasks, contained the narrowest range of MWV lemmas (55). In order to further analyse the range of MWVs used in each task, the top 25 MWVs per task and the contexts in which they appeared were examined. These verbs are presented in Table 4.19 below. MWVs that occurred exclusively and with high frequency in one task are indicated in bold in the table.

Table 4.19 Top 25 MWVs per task

| **Presentation** | **Discussion** | **Interactive task** | **Conversation** |
|---|---|---|---|
| **1. sum up** | 1. grow up | 1. go out | 1. go out |
| 2. find out | 2. go out | 2. go on | 2. grow up |
| 3. move on | 3. go back | 3. come back | 3. go on |
| 4. come back | 4. come back | 4. grow up | 4. come back |
| 5. grow up | 5. wake up | **5. come down** | 5. find out |
| 6. go on | 6. go on | 6. help out | 6. come out |
| 7. point out | 7. come out | 7. keep on | 7. end up |
| 8. wake up | 8. get up | 8. go back | **8. get out** |
| 9. bring up | **9. give out** | 9. take out | **9. hang out** |
| 10. make up | 10. pick up | 10. go down | 10. bring up |
| **11. build up** | 11. keep on | **11. work out** | 11. keep on |
| 12. keep on | 12. show off | 12. come out | 12. take out |
| 13. come out | 13. bring up | 13. pick up | 13. go back |
| 14. carry out | 14. find out | 14. come in | **14. put up** |
| 15. go out | 15. give up | 15. carry on | 15. give up |

| | | | |
|---|---|---|---|
| 16. set up | 16. come in | 16. show off | **16. put down** |
| 17. go down | 17. sit down | **17. take off** | **17. check over** |
| 18. end up | 18. come on | 18. settle down | **18. look up to** |
| **19. go up** | 19. make up | **19. check out** | 19. look up |
| 20. look up | 20. go down | 20. move on | 20. pick up |
| 21. give up | **21. turn off** | **21. clean up** | **21. fit in** |
| **22. set off** | 22. point out | **22. use up** | 22. show off |
| 23. go back | 23. look up | 23. carry out | 23. come on |
| 24. carry on | 24. carry out | **24. pull down** | 24. set up |
| 25. get up | 25. settle down | 25. sit down | 25. help out |

Based on the results from the table, only seven MWVs occurred very frequently in all four tasks (i.e. *come back, grow up, go on, keep on, come out, go out,* and *go back).* Among those shared verbs, several display marked differences in ranking. This is particularly evident between the presentation and every other task. For instance, *go out* is ranked 15 in the presentation but 2 in the discussion, and 1 in the interactive and conversation tasks. Other examples include *go back* (rank 23 in the presentation vs. 3, 8, 13 in the other tasks) and *grow up* (ranked 5 in the presentation vs. 1 in the discussion and 2 in the conversation).

Upon closer inspection, the presentation task is characterised by a large number of MWVs functioning as part of transitions in the context of delivering oral presentations. The verbs *sum up, move on, go on, point out, go back,* and *carry on* helped speakers to summarise, add, and explain information as Examples 27 to 31 illustrate. These verbs tended to be part of larger phraseological structures where the MWVs had a relatively fixed position and often acted as objects of *to*-clauses. Table 4.20 provides an overview of the patterns where MWVs were used as part of transitions in the presentation task along with examples of the MWVs that occurred in those patterns.

27. C: Then they won't there won't be an= so much smog in China any more so to **sum up** I believe we should all help to reduce smog… (Chinese, C1)

28. C: They were as we say the two erm aspects of the everyday life such as war justice wisdom er but also beauty er so er to **sum up** er erm the Greek mythology er was very useful to explain also the supernatural erm events… (Italian, C1)

29. C: When you have the feeling that someone is chasing you that's anxiety and well erm to **sum up** I would like to say that bullying is a problem that we have to take care of… (Spanish, C1)

30. C: The first part of my topic is with pros of adoption in these kind of families and then I would like to **move on** to the potential drawbacks … (Spanish, C1)

31. C: This just creates some visual imagery that overrides the processing of the real image which is the erm actual environment around them so let's **move on** to a more psychological review into the cell phone use… (Chinese, C1)

Table 4.20 MWV use in the presentation task

| Pattern | Examples |
|---|---|
| 1. To + MWV | -To **sum up** I I love so much origami heart (Italian, C1) |
| 2. Would/ 'd like to + MWV + to + noun phrase/verb phrase | -I would like to **move on** to the potential drawbacks (Spanish, C1)<br>-I would like to **go on** to er to describe some of the wildlife in the area (Spanish, C1) |
| 3. Be going to/will/ 'll + MWV + to + noun phrase/verb phrase | -I'm going to **move on** to the disadvantages (Spanish, C1)<br>-I'll **go on** to talk about the controversy around bullfighting (Spanish, C1) |
| 4. Would like/have to + MWV + *that*-clause | -We have to **point out** that er this is a very actual thing (Italian, C1) |
| 5. Will + MWV + verb -ing | -Then I will **go on** talking about the advantages and disadvantages (Chinese, C1) |
| 6. Let + objective personal pronoun + MWV + to + noun phrase | -Let's **move on** to a more psychological review into the cell phone (Chinese, C1) |
| 7. Let + objective personal pronoun + MWV + *that*-clause | -Let me **point out** that this kind of make-up are more advantageous (Italian, C1) |
| 8. Anticipatory *it* + passive MWV + *that*-clause | -It should be **pointed out** that every year more than seventy thousand erm victims reach the shops (Spanish, C2) |

Interestingly, the results showed that some of the MWVs used as transitions in the presentation task also occurred very frequently in the conversation, discussion and interactive phase; however, they did not serve as transitions in those tasks. For instance, *go on* in Examples 33 to 35, which was used to express that an activity continues or takes place, clearly contrasts with *go on* in Example 32, where it functions as a transition in the presentation task to signal that the speaker will move to a new main point.

32. C: Fir-first of all I'd like to explain the reasons and then I will **go on** talking about the advantages and disadvantages… (Chinese, C1, presentation)

33. C: We should er have thi-this habit to erm don't stop ourselves erm front of this diffi-difficulties but we have to **go on** and to to try… (Italian, C1, discussion)

34. C: If we erm fund

    E: mm

    C: erm a military research

    E: mm

    C: the war the wars **go on**

    E: mm

    C: will **go on**… (Italian, C1, conversation)

35. C: Sometimes when you go to university they y-you are in in that world immersed in it and you don't know what's **going on** and what's **going on** outside so… (Spanish, C2, interactive task)

With respect to the verbs that occurred exclusively in one task, results are rather mixed. On the one hand, there was not a strong tendency for those MWVs to be used in connection with specific topics although a few of them were found to be topic-related particularly in the interactive phase and the conversation. Let us remember that those two tasks are based on subject areas previously selected by examiners, who sometimes discuss the same subject area with more than one candidate. In those cases when the same or a similar topic is selected and discussed with more candidates, particular MWVs seem to recur. As a case in point, the PPV *look up to* was mostly used in the conversation task when speakers discussed the topic of role models. In addition, the PV *hang out* was a common choice for speakers who talked about youth behaviour. Two other verbs were consistently used in connection with specific topics: *come down* (interactive task) and *put down* (conversation). These two MWVs differed from other topic-related verbs in the sense that the topics or prompts that the examiners introduced already contained one of those MWVs[17]. In other words, L2 speakers picked up those MWVs from examiners who first introduced them to state the topics to be discussed, and L2 speakers kept using them in their interactions. On the other hand, the analysis of the remaining MWVs occurring in one task only showed that, rather than being repeatedly used to discuss specific topics, L2 speakers chose them to talk about a wide variety of subject areas.

[17] *Put down* was used to answer the question *Is it humane to put animals down? Come down* was used in response to the following prompt: *Some people always seem to rush out and buy the latest smartphone or electronic notebook but I think it's better to wait until prices come down*.

Next, in order to draw finer-grained comparisons of MWV use across tasks, I examined the frequency and range of MWVs in the speech of the most prolific speakers (i.e. those producing five or more MWVs per thousand words in their examinations). Figure 4.4 displays the use of MWVs across tasks by the eight most prolific speakers at the C1-C2 level. In terms of absolute frequency, those eight speakers produced from ten to twenty-one MWVs in their examinations.

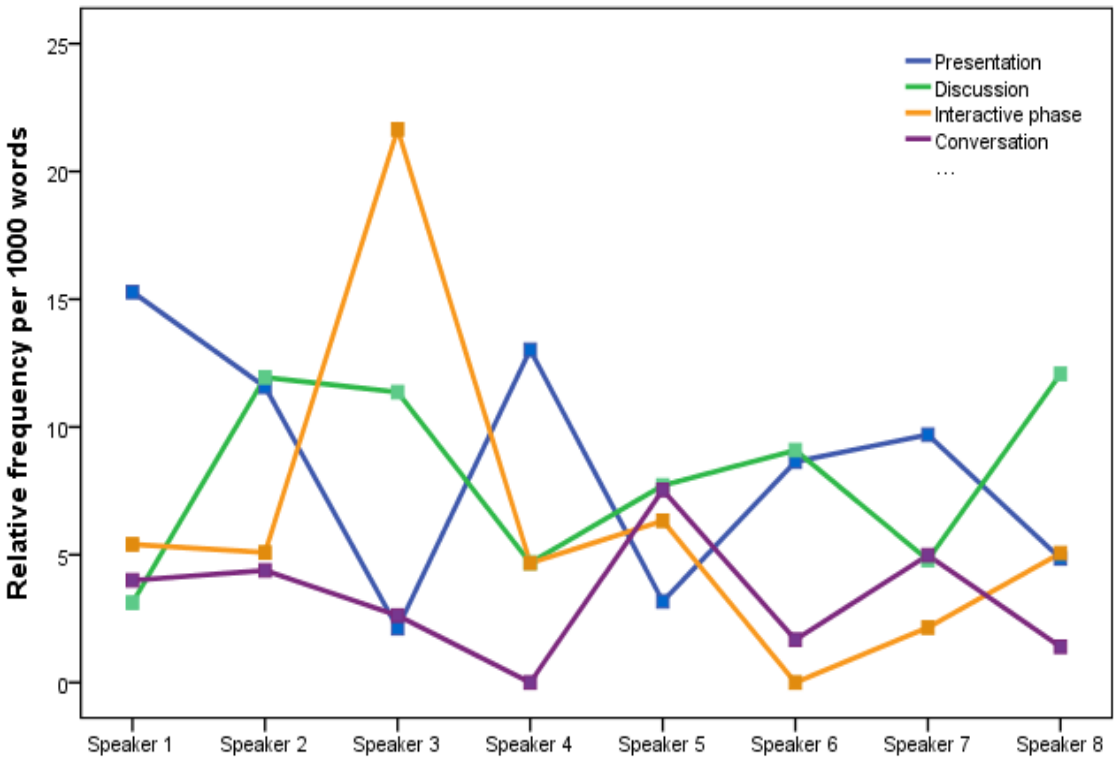


Figure 4.4: Most prolific C1-C2 speakers' use of MWVs across tasks

As is apparent from the figure, there is considerable individual variation across tasks. Despite the fact that no single task elicited the largest number of MWVs in all eight cases, overall the frequency of MWVs in the presentation and the discussion tasks remained relatively high. In contrast, the conversation tended to contain the lowest number of MWVs per thousand words. Speaker 3 is a rather extreme case given that

the presentation elicited the lowest MWV frequency and the interactive task the highest. A closer look at Speaker 3's performance in the interactive task showed that this speaker repeatedly used the same two MWVs (i.e. *help out* and *come back*), thus the high frequency. This speaker's presentation task also contained only two MWVs (i.e. *help out* and *throw up*), yet those were used only once each. The prolific speakers' range of MWVs was assessed in terms of the variety of MWV lemmas that these speakers produced in each task. Findings in this respect showed that the presentation displayed the widest range of MWV lemmas. Speakers in that task produced from two up to six different MWVs. In contrast, the interactive phase elicited the narrowest variety of MWV lemmas with speakers producing no more than two different MWVs in that particular task.

### 4.5. Summary

This chapter presented the findings obtained for the three research questions that the present study aimed to answer. First, it described how MWVs are used by L2 speakers in the TLC. As far as the MWV frequency and coverage are concerned, the study found that overall MWVs occurred 1,397 times pmw. When looking at PV and PPV separately, PVs occurred much more frequently than PPVs. Findings also showed that a small set of twenty MWVs accounts for almost 60 percent of the MWV occurrences in the TLC. The same pattern was found to be true for lexical verbs and particles, i.e. a small number of items made up the majority of MWV occurrences. While the results showed that the most frequent lexical verbs were also the most productive ones, *grow* was found to be an exception because it appeared to be a highly frequent lexical verb that combined with only one particle (i.e. *up*).

Regarding non-canonical MWV forms, a total of twenty were identified and four different patterns of non-canonical MWV use were observed. The first pattern included non-standard MWVs containing a redundant particle (e.g. *restart over*). The second pattern was found in MWVs whose lexical verbs appear to have been confused with the ones in existing MWV forms (e.g. *lay around* vs. *lie around*). The third pattern corresponded to MWVs in which the particle used was close in meaning to the one intended (e.g. *lock off* vs. *lock out*) or functioned as an aspectualiser, thus adding a special meaning to the whole MWV form (e.g. *please up*). The last pattern included combinations of lexical verbs and particles expressing new meanings that, arguably, could not have been expressed using the lexical verbs alone (e.g. *stay back*).

The assessment of polysemy revealed that on average L2 speakers in the TLC produced an average of 1.37 meaning senses per polysemous MWV form. The majority of the most polysemous MWVs are among the most frequent ones identified in this study. Each of the top ten polysemous MWVs expressed between four and nine distinct meaning senses. Despite this fact, an average of one or two meaning senses predominated in the use of each polysemous MWV. In addition, the key senses of the ten most polysemous MWVs in L2 production were found to resemble those reported for the same verbs in L1 production.

The chapter then went on to report the results related to the effect of learner variables in L2 speakers' use of MWVs. The results of the statistical analyses indicated that L2 proficiency was not a statistically significant factor in MWV use, but L1 background was. Chinese speakers appeared to use more MWVs per thousand words than Italian and Spanish speakers. With respect to the range of MWVs, the study revealed five main

findings. These findings were observed not only in relation to L2 proficiency but also L1 background. First, the 25 most frequent MWVs made up the majority of MWV occurrences. A high degree of inter-group variation in the use of those top 25 MWVs was found given that they were very unevenly distributed among proficiency groups and L1 backgrounds. Second, inter-speaker variation within the same L2 group (e.g. B2 speakers) was also observed, with an important number of speakers having no MWVs in their production whereas others used large quantities of them. Third, the set of the most frequent lexical verbs was similar in all proficiency and L1 groups. *Go, come, grow, get,* and *take* appeared in more than half of the MWVs that were produced by each proficiency and L1 group. With the exception of *grow,* all the most frequent lexical verbs were also found to be the most productive and combined with a large number of particles. Fourth, a high degree of overlap in the range of the most frequent MWVs across proficiency levels was found, with 14 out of 25 top MWVs being frequently used by B1, B2, and C1-C2 speakers. In terms of L1 background, however, the degree of overlap observed was much lower, with only eight MWVs being frequently used across the three L1 groups. Five, further inspections of the most frequent MWVs per proficiency group and per L1 background provided evidence of a topic-effect phenomenon whereby particular MWVs recurred in relation to specific topics that the L2 speakers discussed in their examinations.

The last section of the chapter presented the findings regarding the effect of text-specific variables, i.e. task type, on L2 speakers' use of MWVs. Of the four tasks analysed, the monologic presentation elicited most of the MWV occurrences, and it was also the task in which advanced L2 speakers produced more MWVs per thousand words. A large number of MWVs serving as part of transitions in the context of

delivering oral presentations (e.g. *I would like to move on*) were found in the presentation task. These MWVs also occurred very frequently in the other dialogic tasks, yet rather than functioning as transitions, they were found to express various meaning senses. The analysis of high frequency MWVs that occurred exclusively in one of the dialogic tasks revealed that these verb forms were consistently used to talk about specific topics, particularly in the interactive and conversation tasks.

# Chapter 5: Discussion

## 5.1. Introduction

The aim of the present study was to examine MWVs, i.e. PVs and PPVs, in L2 spoken communication using corpus methods. In particular, the study looked at L2 MWV use from three different angles, taking into consideration i) the MWVs themselves and their overall frequency, semantic behaviour and usage patterns in the corpus, ii) the learners who used those MWVs, and iii) the tasks in which the MWVs were produced. In this chapter, the findings are discussed in light of those three angles, and possible explanations for the patterns found are provided. Section 5.2 first looks at the frequency, coverage, and meaning senses of MWVs in the TLC relative to the ones previously reported in the contexts of L1 and L2 spoken and written communication. It also discusses the patterns observed in the use of non-canonical MWV forms. Next, the relationship between MWV use and learner variables (i.e. L2 proficiency and L1 background) is addressed in section 5.3. Section 5.4 is devoted to the effect of task type on the frequency and range of MWVs produced by advanced L2 speakers.

## 5.2. Use of MWVs by L2 speakers

### 5.2.1. Frequency of MWVs in the TLC

In terms of overall frequency, this study found that MWVs (i.e. both PVs and PPVs) occurred 1,421 times per million words (pmw) in the TLC. However, comparing that result to the frequencies reported in earlier studies is problematic given that, unlike the present study, previous research has tended to focus exclusively on PVs (e.g. Gardner

& Davies, 2007; González, 2010; Chen, 2013, 2018; Deshors, 2016). Therefore, I will first discuss the findings related to PV frequency and then refer to PPVs.

The results showed that PVs occurred 1,397 times pmw. It should be noted i) that this rate of use was similar even when the data were analysed with respect to speakers' L2 proficiency (see section 4.3.1.1) and ii) that large inter-speaker variation was found in L2 MWV production (see section 5.3.1). The frequency of PV occurrence in the TLC appears to differ considerably from the one observed by Weirszycka (2013) and Gilquin (2015b), the only two previous corpus-based studies on spoken L2 communication that report frequencies considering all MWVs found in the corpora that they used rather than just selected MWV forms (e.g. Gilquin, 2011; Märzinger, 2013). In her analysis of PVs in the Polish component of the LINDSEI corpus, Weirszycka (2013) found that L2 speakers produced 2,367 PVs pmw. A very similar PV frequency (i.e. 2,472 PVs pmw) was reported by Gilquin (2015b)[18], whose constructional and collostructional study of PVs also drew on the LINDSEI corpus, among other corpora, but focused on French L2 production. Regarding PVs in written L2 communication, the differences between the PV frequency found and those reported previously are even more considerable, with studies indicating that PVs occurred from 3,000 up to 6,500 times pmw in L2 academic writing (e.g. Waibel, 2008; Gilquin, 2015b; Chen, 2013, 2018; Deshors, 2016).

In corpus-based studies looking at PV production in L1 English and both spoken and written registers, PVs have also been found to occur more frequently than in the TLC.

[18] Gilquin (2015b) analysed the frequency of PVs at three hierarchical levels associated with the PV construction. The PV frequency used for comparison purposes in the present study is the one Gilquin reports at the higher level or 'superconstruction' as this corresponds to the overall frequency of PVs in the corpus.

While Biber et al. (1999) found that PVs occurred 2,000 times pmw in the LSWE corpus, Gardner & Davies (2007) and Liu (2011) reported PV frequencies of 5,189.23 and 3,210 pmw respectively in the BNC. More recently, drawing on very extensive evidence that combined written and spoken data from COCA, the BNC, the Corpus of Global Web-based English US (GloWbe US) and the Corpus of Global Web-based English UK (GloWbe UK) (Davies, 2013), Gardner & Davies (2018) showed that PVs occurred 6,512 times pmw, a frequency that is almost five times the one found in the present study. It is worth pointing out, however, that the wide range of PV occurrence reported in these studies (i.e. from approximately 2k to over 6.5k) shows that there were also considerable differences in the PV frequency found in studies on L1 communication. These differences could most likely be attributed to variations in the composition and nature of the interaction represented in these corpora (e.g. spoken and written mode, different genres and registers).

With respect to the reasons for the differences between the frequency of PVs in the TLC and those identified previously, there are two possible interconnected explanations: one is related to the nature of communication represented in the TLC and the other to the methodological differences in the identification of PVs across studies.

While the results of this study show that PVs occurred less frequently in the TLC than in other corpora representing L1 and L2 spoken and/or written communication, they are still in line with previous observations regarding the lower PV representation in L2 speech than in writing (e.g. Gilquin, 2015b; Deshors, 2016). De Cock (2005) argues that L2 speakers are 'stylistically deficient' when it comes to producing MWVs, which means that they might use "more PVs in formal writing than in informal speech" (p.

LS16). This lower awareness of register constraints mentioned by De Cock might partially explain the large differences between the PV frequency found in this study and those reported in L2 writing. In the same vein, it is possible that the examination context in which the L2 speakers in the TLC had to interact might have led them to opt for more formal vocabulary, thus avoiding PVs because they considered them informal or inappropriate to communicate in an institutional setting where they were being evaluated. It is also likely that in the examination, a high-stakes context, the L2 speakers opted to avoid what they may have considered a more risky structure and relied on safer lexical choices (see section 5.3.2).

Another plausible explanation for the higher PV frequency in written L2 corpus-based studies relates to the fact that they have investigated speakers' production mainly in one written genre, that is, argumentative essays. As also demonstrated by studies on L1 communication, it is likely that a specific genre may play a major role in PV frequency. With respect to the rate of PVs found in L2 writing, not only were the essays written under planned conditions, but they were produced by advanced university students. It could be argued that a context where the L2 users had time to think what they wanted to write gave them an advantage to carefully select PVs as opposed to a conversation where L2 speakers had to use PVs spontaneously. This explanation seems plausible as this study found that advanced L2 speakers produced most PVs in the task that provided them with planning time (i.e. the presentation) (see section 5.4.1). It is also quite possible that advanced university students were trained in the written use of academic PVs and other multi-word units, which might have resulted in a higher PV frequency.

The higher frequency of PV occurrence in previous spoken and written corpus-based studies might also be related to two methodological factors. First, various ways have been adopted to define and operationalise the term PV. For example, in Waibel's (2008) and Ke's (2013) studies, PPVs were subsumed under the PV category. In other words, both studies regarded PPVs as PVs, which might have resulted in a higher PV frequency. Second, the approach followed to manually clean the data so that false positives could be excluded is likely to have varied from one study to another. As a consequence, it is possible that some researchers were more flexible than others regarding the verbs that they were counting as PVs. This is reflected in the fact that studies looking at PVs in the same corpora and adopting the same definition of PV report incongruent PV frequencies (e.g. Gilquin, 2015b and Deshors, 2016 for the French component of the ICLE corpus; Weirszycka, 2013 & Gilquin, 2015b for LOCNEC). Because of the similarity between PVs and other MWVs (e.g. prepositional verbs, see section 2.2.2), manual scrutiny of the results is essential. Manual analysis is indeed mentioned as one of the methodological steps followed, but few studies (e.g. Märzinger, 2013) provide an account of how it was actually carried out. Rather than an oversight, the lack of a description of the coding scheme used to manually filter the search results and its application is likely to have been the result of the space limitations in previous studies. In the present study, the description of the coding scheme developed for the purpose of manual data analysis is provided in section 3.10.3. Moreover, given the inherent subjective nature involved in distinguishing PVs from similar MWV forms, the manual analysis was followed by a double-coding process, and the level of inter-rater agreement was established and reported (see section 3.10.4).

As far as PPVs are concerned, little is known about their overall frequency in speech. Biber et al. (1999) state that PPVs occurred 300 times pmw in the LSWE corpus, a frequency that is considerably higher than the one reported in the present study (i.e. 24.2 times pmw). It is not surprising that the frequency of PPVs in the TLC is much lower than that of PVs. This also reflects the pattern that Biber et al. (1999) reported in the LSWE corpus. Also, even in studies that have grouped PPVs and PVs together, PVs tend to outnumber PPVs as the frequency lists of all the MWVs found in those studies show (e.g. Märzinger, 2013; Waibel, 2008; Ke, 2013). The low representation of PPVs in L2 speech can be attributed to the overall paucity of PPVs in English (Claridge, 2000). From a semantic perspective, the fact that PPVs are usually idiomatic and rather opaque in meaning (Claridge, 2000; Waibel, 2008, Rodríguez-Puente, 2019) might have also led L2 speakers to avoid their use in conversation, which in turn explains their infrequent use.

### 5.2.2. MWV coverage, lexical verb and particle productivity

In this study, MWV coverage refers to the percentage of MWV occurrences that each MWV lemma accounts for. Productivity is understood as the ability of an element (i.e. a lexical verb or particle) to combine with different particles (in the case of verbs) or verbs (in the case of particles) in order to form lexically distinct MWVs. Put differently, the most productive lexical verbs are those that appear next to many different particles in a large number of distinct MWVs. Similarly, the most productive particles correspond to those that combine with many different lexical verbs to form a large number of distinct MWVs.

As far as the coverage of MWVs is concerned, the present study brought to light that a small group of 20 MWVs accounts for 60 percent of the MWV occurrences in the TLC. This pattern was also observed in the frequencies of lexical verbs and adverbial particles, i.e. that a small number of high-frequency items make up the majority of MWV occurrences. On the one hand, these findings are in line with what Gardner & Davies (2007), Liu (2011) and Lee (2015) found for native speaker production in the BNC, COCA, and MICASE corpora respectively (see section 2.3.1.1). On the other hand, they provide evidence that MWVs in L2 speech also follow a Zipfian distribution (Zipf, 1935), which has been described as a common pattern in vocabulary learning (Ellis, 2012; Nation, 2012; Webb & Nation, 2017; Wulff, 2019). What is more, those findings show that Zipf's law seems to operate not only at the single lexical verb and particle levels but also at the phraseological MWV level, thus supporting previous claims that Zipf's law is more evident in the distribution of multi-word units than in that of single words (Williams et al, 2015; Christiansen & Arnon, 2017).

At the MWV level, the verbs *go out*, *grow up* and *come back* were overwhelmingly more frequent than the rest of MWVs and found in a large number of the MWV occurrences. With the exception of *grow up* which is further discussed in section 5.3.5, these verbs usually rank among the ten most frequent MWV lemmas in spoken and written L1 English as previous research has shown (Gardner & Davies, 2007; Liu, 2011; Lee, 2015), and they also appear to be very frequently used by L2 speakers (Märzinger, 2013; Gilquin, 2015b). One reason why *go out* and *come back* make up a large number of MWV occurrences could be tied to the lexical verbs that they contain. Both *go* and *come* are highly frequent lexical verbs (Altenberg & Granger, 2001; Leech et al., 2001; Gouverneur, 2008), which might make them salient and more likely to be learned than

less frequent, non-salient lexical verbs (Ellis, 2006). In fact, Kaszubski (2000) argues that L2 users tend to overuse those verbs. Another plausible explanation for the high frequency of *go out* and *come back* could be a semantic one. Both *go out* and *come back* tend to be less opaque in meaning than other MWVs (e.g. *give up, make up*), which could make them easier to use by L2 speakers.

From the perspective of lexical verbs, the study provided evidence of a frequency-productivity phenomenon where the most frequent lexical verbs (i.e. *go, come, get,* and *take*) are also the most prolific ones; in other words, they occurred together with a wide variety of particles to form MWVs. These lexical verbs have been previously identified as highly frequent (Altenberg & Granger, 2001; Nesselhauf, 2005) and represent some of the most productive verbs in MWV formation in native production across English varieties and registers (Liu, 2011). In addition, their high degree of polysemy (Biber et al., 1999) increases their frequency in different contexts. The case of *grow* is an interesting exception as this verb was particularly frequent, yet it was one of the least productive ones. Possible reasons for the high frequency of *grow* as well as of the combination *grow up* are discussed in section 5.3.5.

With respect to particles, findings showed that learners used a relatively wide range of particles. It is worth noting that the frequency rank of particles found in this study broadly resembles the one reported for native-speaker use of MWVs (Biber et al., 1999; Gardner & Davies, 2007). As expected on the basis of historical accounts of particle frequency (e.g. Hiltunen, 1994), *up* and *out* were both the most frequent and most productive particles. The high frequency of both particles can be attributed to two reasons. First, *up* and *out* are more likely to act as particles than as any other

grammatical forms, namely, prepositions (Dehé, 2002; Huddleston & Pullum, 2002; Gardner & Davies, 2007). Second, their multiple meanings, both concrete and abstract, allow them to appear next to many different lexical verbs to express a wide variety of meanings (Side, 1990; Lindner, 1983; González, 2010; Lindstromberg, 2010).

### 5.2.3. Non-canonical MWV forms

As was mentioned in the results chapter (section 4.2.3), in this study non-canonical forms correspond to combinations of lexical verbs and particles that are not recorded in the MWV dictionaries consulted. This leaves aside cases where, for example, a MWV was used (e.g. *set out an aim for oneself*), yet a simple verb could have been more appropriate (e.g. *set an aim for oneself*) (Nesselhauf, 2005). It is worth pointing out that two possible reasons why non-canonical MWVs occurred in L2 production might be erroneous MWV use and creative MWV use. Some non-standard MWVs appeared to be the result of erroneous use of an attested MWV (i.e. the target verb seems to correspond to a MWV form different from the one produced) (e.g. *sack up* instead of *stack up*) while others appeared to be instances of lexical creativity (i.e. the target MWV is the actual non-canonical form produced) (e.g. *become out*). In total, 20 non-canonical forms were found in the TLC; the majority of those corresponded to one-off occurrences. This finding is somewhat surprising considering the difficulty that MWVs pose to L2 speakers (Siyanova & Schmitt, 2007) and the usual presence of deviations from established forms that characterise learner language (van Rooy, 2011). With respect to the distribution of non-canonical MWVs across proficiency levels, the study showed that the more proficient speakers (i.e. from B2 and C1-C2 bands) were responsible for the majority of non-canonical MWVs. Because of their high proficiency, those speakers were expected to produce fewer non-canonical MWVs.

However, the presence of these forms in advanced speech could partly be explained by the fact that advanced speakers have a better command of the language and thus can risk creating verb-particle combinations as opposed to beginner B1 speakers who may have a more limited repertoire of linguistic resources.

The non-canonical forms found broadly fall into four patterns of use. The first pattern concerns the presence of a redundant particle as in *restart over* (seven cases). The second pattern involves the use of a lexical verb whose pronunciation is close to one of other similar verbs in attested MWVs (e.g. *lay around* vs. *lie around*) (two cases). The third pattern was identified in forms where the particle adds a special meaning to the whole MWV (e.g. *please up*) (eight cases), while a fourth pattern was observed in combinations of verbs and particles that helped learners express new meanings (e.g. *become out*) (three cases). Regarding the first pattern, non-canonical MWV forms with a redundant particle are a common phenomenon in learner language (e.g. *rise up, complete up, increase up*) (Gilquin, 2011, 2015a; Schneider & Zipp, 2013; Schneider & Gilquin, 2016) as well as in L1 English varieties (e.g. *boost up, double up, finish up, finish off, fall down, sit down*) (Hampe, 2002; Schneider, 2004; Thim, 2012; Zipp & Bernaisch, 2012). In fact, Hampe (2002) notes that MWVs with redundant particles in L1 speech tend to function as indices of the emotional involvement of speakers and help them emphasise the actions expressed by the lexical verbs in the MWV forms. Given the spoken nature of the data in the TLC, it is of course possible to argue that some non-canonical forms with additional particles were merely the result of speakers' attempts to repeat an idea. For instance, the candidate in Example 36 might have just been trying to repeat the verb *download* rather than intending to use *down* as a particle. In order to check this possibility, the audio files for the corpus would need to be consulted, which

was beyond the scope of this research. It should be noted, however, that even having access to the audio evidence might prove insufficient to resolve the issue.

36. C: I 've seen a lot of foreign websites and I really like them and about the songs that I wanted to

E: mm

C: **download it down** they

E: yeah

C: s= they saisd that I needed money like several dollars… (Chinese, C1)

As regards the second pattern, non-canonical MWVs in which the lexical verb seems to have been confused with another verb with similar pronunciation (e.g. *lay around* vs. *lie around, sack up* vs. *stack up*) possibly occurred due to three reasons. First, these MWVs are likely to reflect an L2 pronunciation issue in which the L2 speakers seemed to be unsure of how to appropriately pronounce the lexical verb in an attested MWV and thus opted for using a similar-sounding verb that they were familiar with. Second, the use of these MWV forms might be related to a semantic issue in which L2 speakers possibly confused the meaning of two semantically similar verbs (e.g. *lay* and *lie*). Third, we could speculate that the specific case of *sack up* vs. *stack up* might have just been the result of a transcription issue in which the lexical verb was accidentally misspelled by transcribers. Having access to audio files could possibly be used to clarify whether this particular non-canonical MWV was the result of a transcription issue.

Instances of non-standard MWVs categorised under patterns three and four deserve special discussion as they involve the creation of what L2 speakers think are plausible

lexical verb and particle combinations. As mentioned above, patterns three and four include MWV forms where the speakers either added a particle that contributed a perfective meaning to the lexical verb (e.g. *please up*) or combined verbs and particles to express new meanings that arguably could not have been expressed using the lexical verbs alone (e.g. *become out*). Although the meaning of some of those combinations is difficult to interpret as Example 37 illustrates, others seem to make sense in the context where they are used. For instance, the particle *back* in Example 38 specifies that a change has to take place so that the person returns to the way he looked before. In Example 39, the speaker is advising the examiner about lending money to a friend. *Around* in such example complements the action of spending by adding a notion of carelessness, one of the common metaphorical meanings of *around* (Lindstromberg, 2010) that the verb alone does not convey.

37. C: I don't think I would have an affair

    E: uhu

    C: but er if sometimes would emotionally sometimes

    E: yes

    C: they go

    E: mm

    C: beyond your your to **catch** them **back** I mean and I'll tr= I'll try to stop them before they go

    E: mm

    C: too far… (Italian, B2)

38. C: So you maybe you want er help him to change to **change back** to **change back** to come back at the previous er look… (Italian, B2)

39. C: You said that you're encouraging a

    E: mm

    C: a habit maybe if you don't give him moment he won't you know

    E: yeah that's a bad habit

    C: the feeling

    E: mm

    C: maybe he's you know **spending** it **around**… (Spanish, B2)

Based on Examples 37-39, it is likely that some of the non-canonical MWVs categorised as patterns three and four were produced as a result of L1 transfer, which has been considered a key factor directly influencing L2 lexical creativity (Dewaele, 1998; James, 1998). In the examples above, for instance, *spend [money] around* could have been the product of a literal translation of the Spanish phrase *gastar [el dinero] por ahí*, which refers to the action of spending money without a purpose in mind. We can also speculate that the Italian speaker in Example 37 created the MWV *catch back* in analogy with a similar-sounding verb form in Italian (i.e. *cacciare indietro, chase back*), with the pronunciation of *cacciare* closely resembling that of *catch* in English.

The fact that L2 speakers can add particles and put together verbs and particles to help them in conversation as shown in the previous examples provides strong evidence of speakers' relatively high level of phraseological knowledge and testifies to their potential to create MWVs. It can safely be assumed that in order to add particles or

combine particles and verbs to express an idea, one has to be aware that those combinations exist in English in the first place. In addition, one should be familiar with the meaning of both particles and verbs to be able to put them together in context. As Waibel (2008) observes in MWV use by advanced German speakers of English, L2 speakers should be "familiar enough with existing English phrasal verbs to form new ones or use combinations that are unusual also in native English" (p. 150). Although Waibel's observation was made regarding advanced German L2 speakers, it is likely to be applicable to other advanced L2 speakers as well whose generally large linguistic repertoires may facilitate their engagement in creative language use (Ellis, 2012). In this study, it is actually the L2 speakers at the higher levels of proficiency (i.e. B2 and C1-C2 speakers) that produced most of the non-canonical MWVs conforming to patterns three and four. As previous SLA research has shown, lexical creativity is exemplar-based (Myles, Mitchell & Hooper, 1999; Ellis, 2012; Eskildsen, 2016). In other words, speakers base their creative uses of a target form on previous examples of this form that they are familiar with and on the abstraction of regularities within those examples. In the context of MWV formation, this means that L2 speakers use their previous MWV knowledge as the basis material to generate new MWV forms even if these forms are considered unconventional or deviant.

### 5.2.4. Polysemy and L2 semantic knowledge

This study showed that the majority of the most frequent MWVs are also the ones displaying the largest number of meaning senses, thus supporting the notion that there is a strong correlation between frequency and polysemy in vocabulary use (Crossley, Salsbury & McNamara, 2010; Boers, 2013; Larsen-Freeman, 2002). These results complement findings about the relationship between frequency and polysemy found in

single lexical verbs (e.g. Viberg, 2002) by showing that such relationship also occurs in multiple-word verbs.

Moreover, the findings about the occurrence of meaning senses in the TLC are consistent with previous investigations of MWV polysemy showing that only one or two key senses dominate the use of a MWV regardless of how polysemous such verb is (Garnier & Schmitt, 2015). The dominant key senses are not only the most frequent ones but also tend to be the core meanings of MWVs (e.g. *come back: return to a place, wake up: cease to sleep*). The fact that one or two key senses dominate the use of a polysemous MWV can be explained in terms of the different degrees of salience and the frequency variations that characterise the multiple meaning senses of a single lexical item.

With respect to salience, not all the meanings of a polysemous word or phrase are equally accessible to language users "because we ascribe greater cognitive priority in our mental lexicon to some meanings over the rest" (Kecskes, 2006, p.221). More salient meaning senses will thus be those that are more familiar to speakers based on their prior knowledge, experience and encounters with those meanings (Giora, 2003). While L2 speakers may be aware of the existence of multiple meanings for a single MWV, they may give priority only to those senses that they have used in previous similar and typical contexts. Regarding frequency, cognitive and corpus-based approaches to lexical semantics (e.g. Vespoor & Lowie, 2003; Gries, 2015) have shown that not all the meaning senses of a polysemous lexical item are equally frequent, and often a rather limited number of senses tend to account for the majority of instances in which such lexical item is used. The present study found that this pattern also holds true

for MWVs in the TLC (i.e. a few senses accounted for most instances where a polysemous MWV was used) and adds to previous findings by showing that the dominant key senses of highly polysemous MWVs in L2 speech tend to match those reported for the same verbs in L1 speech (Liu & Myers, 2018) (see section 4.2.4). What is more, L2 speakers' knowledge of the frequent key senses appears to support previous claims (Schmitt & Redwood, 2011; Garnier & Schmitt, 2016) indicating that the more frequent a meaning sense is, the better L2 speakers appear to know this meaning.

Overall, the number and variety of meaning senses found in this study are indicative of a wide range of semantic knowledge among the L2 speakers in the TLC as their repertoire of meaning senses was often not limited to the most frequent ones but also included less common senses. In fact, the total number of meaning senses (key and non-key) associated with some of the most frequent MWVs in this study was much greater than that reported previously for L2 production (e.g. Chen, 2018). For instance, while L2 speakers in this study used *come back* in its dominant, key sense (i.e. return to a place, Example 40), they also used that verb in a variety of meaning senses as Examples 41 to 44 illustrate. The meanings of *come back* in each of the examples are provided in brackets.

40. C: She has lunch at home because she follows the lessons the courses in the morning then she **comes back** home for lunch at one o'clock… (Italian, C1) (return to a place)

41. C: I think that ev-everything er is **coming back** I mean erm you can see this erm kind of tendency or this trend of

E: mm

C: on music

E: mm

C: you can see well you can listen to some songs that we we used to we used to enjoy when we were er younger… (Spanish, C1)

(become fashionable again)

42. C: There is a threat that that political party the one I was telling you before that it was er ruling for over sixty years is **coming back**… (Spanish, B2)

(be restored)

43. C: He could take part in er different competitions any more

E: mm

C: er fortunately he just needed to rest for a while and

E: uhu

C: now he **came back** and

E: uhu

C: tried harder to recover… (Chinese, B1)

(return successfully to prominence or fame)

44. C: When people get used to language and language you know er like

E: mm

C: er vice versa it c= it **comes back** to our

E: sure

C: human minds and er in= er in-influence quickly our er human mind… (Chinese, C2)

(return to memory)

It should be noted that there might be some variation in the number of meaning senses listed in previous research and this study because some related MWV meanings can be grouped or split in different ways. For example, it is possible to group the meanings of *come back* in Examples 40, 43 and 44 together under the meaning sense 'return' rather than to split them into three meaning senses the way it was done in this study.

## 5.3. The role of learner variables

One of the aims of the present study was to determine the effect of learner variables (i.e. L2 proficiency and L1 background) on L2 speakers' use of MWVs. MWV use was operationalised in terms of the frequency of MWV occurrence and the range of MWVs across three proficiency levels (B1, B2 and C1-C2) and three L1 backgrounds (Chinese, Italian, and Spanish). Given that the statistical analyses conducted indicated that the interaction between the effects of L2 proficiency and L1 background was not statistically significant, the results will be discussed in light of the individual main effects that these two variables had on the frequency and range of MWVs.

### 5.3.1. L2 proficiency and MWV frequency

The findings related to the frequency of MWV occurrence indicated that there is not a clear linear relationship between L2 proficiency and MWV use. The statistical analysis of L2 proficiency did not provide evidence of a significant main effect of this variable

on MWV frequency. This means that an increase in L2 proficiency does not necessarily translate into a higher frequency of MWVs. Although a very modest increase in the frequency of MWVs could still be observed with advanced speakers appearing to use more MWVs per 1000 words than the other two groups (see section 4.3.1.1), differences in the frequency of MWVs between B1, B2 and C1-C2 levels appear to be rather small.

The fact that gains in L2 proficiency do not go hand in hand with a more frequent use of MWVs has been observed in previous studies comparing MWV frequency in intermediate and advanced L2 speakers' production (e.g. Becker, 2014). In fact, Hulstijn & Marchena (1989) argue that both intermediate and advanced L2 speakers are equally likely to avoid MWVs and often prefer semantically equivalent one-word verbs. Despite the existence of these studies that did not find a link between L2 proficiency and MWV frequency, an important body of evidence mainly coming from elicitation studies has shown the reverse tendency, i.e. that L2 proficiency was indeed a key factor mediating MWV use and that higher proficiency correlated with higher MWV frequency (e.g. Dagut & Laufer, 1985; Yorio, 1989; Liao & Fukuya, 2004; Siyanova & Schmitt, 2007).

One of the reasons why the relationship between L2 proficiency and MWV frequency remains unclear is likely due to the fact that the connection between these two is a relatively complex issue. It has been previously reported that the productive use of MWVs like many other types of MWUs proves to be challenging for L2 learners at various stages of language development (Cowie & Mackin, 1993; González, 2010; Gardnier & Schmitt, 2015). In fact, the difficulty that MWVs present to L2 users regardless of their proficiency level has also been found to be true for many other

formulaic units. As has been previously argued (Wray, 2002; Ellis, 2012), having a better command of the language is not necessarily associated with a more frequent use of formulaic sequences. Even advanced speakers who are able to amass sufficient L2 vocabulary including a wide repertoire of single words can show poor formulaic knowledge (Wray, 2002). Therefore, the low occurrence of MWVs in the spoken production of both low and high proficiency L2 speakers might actually be reflective of the usage patterns of formulaic sequences in L2 contexts where there does not seem to be a direct relationship between L2 proficiency and formulaic language use.

The inconsistency in previous research findings regarding the link between increasing L2 proficiency and higher MWV frequency can also be attributed to the various ways in which L2 proficiency has been operationalised. For example, all the following factors have been taken as a proxy for L2 proficiency in previous studies: length of stay in an English-speaking country (Irujo, 1993; Weirszycka, 2013), amount of L2 instruction (Hulstijn & Marchena, 1989; Zhang & Wen, 2019), researchers' subjective appraisal of speakers' proficiency level (Dagut & Laufer, 1985), enrolment in an intensive English program (Becker, 2014), and institutional status (e.g. 4$^{th}$-year undergraduate students of English considered to be advanced L2 speakers) (Siyanova & Schmitt, 2007). In other cases, proficiency levels have been externally estimated or assigned based on in-house language examinations or vocabulary size tests, without studies providing further details or descriptions of the individual proficiency bands. Therefore, it is difficult not only to compare the findings of this study to those reported in previous investigations on the role of L2 proficiency in MWV frequency, but also to ascertain the significant effect of L2 proficiency on MWV use due to the multiple ways in which the construct of proficiency has been operationalised (Chen & Smakman, 2016).

Finally, the unclear relationship between L2 proficiency and MWV frequency could also be related to the role of individual variation in MWV use. In L2 studies, proficiency is considered "a major source of variation" (Gablasova, Brezina & McEnery, 2017). What is more, there can still be considerable variation between speakers from the same proficiency group (e.g. B1 or B2 speakers) (Ortega & Byrnes, 2008; Vyatkina, 2012; Ädel, 2015; Callies, 2015). From the perspective of MWV production, this means that speakers at the same proficiency level do not use MWVs homogenously and might even display idiosyncratic patterns of MWV frequency. As shown in the results of the present study (see section 4.3.1.2), MWVs were absent in the production of many B1, B2, and C1-C2 speakers while others used large quantities of them. These differences in the frequency of MWVs observed at all levels of proficiency suggest that the relationship between proficiency and MWV frequency is a complex issue rather than a straightforward one and that not all speakers can be expected to produce larger numbers of MWVs as their proficiency level increases.

### 5.3.2. L2 proficiency and the range of MWVs

Even though L2 proficiency was found to be a statistically non-significant factor in the frequency of occurrence of MWVs, the patterns of MWV use observed across the proficiency groups merit special discussion. As described in section 4.3, the range of MWVs was analysed in terms of the most frequent MWVs across the three proficiency groups, the most common lexical verbs functioning in MWVs, and the specific set of the most frequent MWVs per proficiency band. Considering those three aspects, the present study brought to light four important findings related to the range of MWVs. These four findings will now be summarised and discussed.

First, the coverage patterns of MWVs and lexical verbs described in section 5.2.2 about overall MWV frequencies also hold true for MWV use at all proficiency levels. A small group of MWV forms accounted for the majority of MWV occurrences in B1, B2 and C1-C2 speech. The same pattern was observed in the use of lexical verbs, i.e. only five lexical verbs (*go, come, grow, get,* and *take*) made up the largest proportion of MWVs. The explanations for those patterns that are provided in section 5.2.2 and that relate to frequency and salience of lexical verbs also apply to the context of L2 proficiency and MWV use. The possible reasons for the high frequency of the lexical verb *grow* will be discussed in section 5.3.5.

Second, even within the same proficiency band, there was a high degree of inter-speaker variation reflected not only in the frequency of MWVs but also in the variety of MWV lemmas. While some L2 speakers from all proficiency levels used eight or more MWVs per thousand words, many others did not produce any. Even in cases where the speakers produced several MWV forms, the study showed that a high MWV frequency was not equivalent to a wide variety of MWV lemmas. This is expected considering that a speaker can use the same MWV form multiple times, thus increasing the frequency of MWV occurrence. Inter-speaker variation is a common phenomenon in L2 use as it has been shown that patterns found at the group level might not reflect those at the individual speaker level (Lowie & Verspoor, 2015; Verspoor et al., 2017). However, large differences in the number of MWVs produced by speakers within the same proficiency band and who engaged in the same tasks and even discussed the same topics raise questions as to which linguistic resources the L2 speakers used instead of MWVs. One possibility is that speakers opted for semantically equivalent one-word verbs, a strategy that has been observed in previous studies on L2 MWV use (e.g. Yorio, 1989;

Siyanova & Schmitt, 2007). For instance, while *pick up* was frequently produced by B2 speakers in the conversation task to talk about the topic of pollution and recycling (Example 45), the one-word equivalent *collect* was also found in the same context (Example 46). It is also possible that, instead of MWVs, speakers produced the lexical verb of a contextually appropriate MWV but omitted its particle (Example 47). Still other speakers might have attempted to paraphrase the meaning of a MWV like the speaker in Example 48 appears to have done by describing a meaning that seems to be conveyed by the MWV *dress up*.

45. E: Does the government actually recycle well erm do you think?

    C: depends where you live

    E: mm okay

    C: in Mexico some neighbourhoods er the the government is in charge to **pick up** the garbage…

    (Spanish, B2)

46. E: Where was the rubbish? in the sea or on land

    C: everything e-e-everywhere everywhere

    E: really yeah

    C: in the streets er in erm near the house er

    E: why is that?

    C: because the government erm er **collect** all the rubbish in north Italy and er send it to Naples...

    (Italian, B2)

47. C: The the persons that are recol= that **pick** the garbage

    E: mm

    C: used to put all the garbage in one only truck…

    (Spanish, B2)

48. C: If he had he has to talk with someone

    E: mm

    C: with clients or customer

    E: mm

    C: they have to be **wear er formal clothes**…

    (Spanish, B1)

Third, the range of the most common MWVs in B1, B2, and C1-C2 speech remains largely unchanged, with speakers from all proficiency groups recurrently using a core set of 14 MWVs (see Table 4.10). Low and high proficiency L2 speakers' reliance on this particular core set might be partially attributed to the functionality of the MWVs in it. Closer analyses of this core set showed that most of the MWVs are highly polysemous (e.g. *pick up, get up, go back, come on, come out*), and thus extremely functional because they can be used in a wide range of contexts. As research shows, more frequent and functional sequences are easier to acquire as opposed to less functional ones which are used less frequently by L2 learners (Wood, 2015). Another likely explanation for the overlap in the range of MWVs per proficiency band might relate to the safety that those high frequency, polysemous MWVs offer to the L2 speakers (Granger, 1998b, 2019; Kaszubski, 2000). Echoing Hasselgren's (1994) concept of the 'lexical teddy bear' (i.e. safe words or phrases that L2 users tend to stick

to and feel comfortable using), the fact that the most frequent MWV forms are almost the same in B1, B2 and C1-C2 speech might be indicative of a MWV teddy bear phenomenon whereby speakers, regardless of their proficiency level, employ the same group of MWVs, which they find safe and widely usable across topics and communication contexts. Even speakers who are sufficiently advanced to risk using other MWVs opted for these well-known forms in spoken communication.

The last and fourth finding related to the range of MWVs per proficiency band is that the topics discussed in the tasks seem to have a strong influence on the use of MWVs. This finding is in accordance with what earlier studies (Hinkel, 2009; Tracy-Ventura & Myles, 2015; Khabbazbashi 2017; Yoon, 2017; Caines & Buttery, 2018) observed regarding certain task topics encouraging the use of specific lexico-syntactic structures, which in turn may result in a higher frequency of these structures (Brezina, 2018b; Gablasova et al., forthcoming). In the TLC, a large number of MWVs were repeatedly chosen by B1, B2 and C1-C2 speakers to talk about the same topics, particularly in the conversation task. Based on the patterns observed, there could be two possible reasons why task topics appear to activate the use of particular MWVs in the production of both low and high proficiency speakers. First, some of the concepts associated with certain topics (e.g. fashion, childhood) are typically expressed by using MWVs (e.g. *dress up, grow up*). In other words, some MWVs are likely to come up naturally in conversations about those topics, and therefore they will be more frequent whenever those topics are discussed. Second, it is also possible that in their previous learning experiences L2 speakers were taught some MWVs as topical vocabulary that they could use to discuss specific subject areas, which they later encountered in the conversation task (e.g. routines, family, environment, social issues). As a result, if these subject areas come up

often in the corpus, the specific MWVs that the speakers learned to use with them will tend to recur. It could be argued that in these cases the topics act as a stimulus and help to activate previously known MWVs.

### 5.3.3. L1 background and MWV frequency

From the three L1 backgrounds analysed, Chinese speakers appeared to use more MWVs per thousand words than Italian and Spanish speakers. The MWV frequency in the speech of Chinese speakers closely resembles the one reported for L1 speakers in spoken communication (Biber et al., 1999). Before elaborating on possible reasons for the frequent use of MWVs by Chinese speakers, let us mention that the higher MWV frequency in Chinese L2 speech is incongruent with previous findings on MWV use and L1-L2 structural distance (see section 2.3.2.1). A large body of elicitation and corpus-based research on MWVs (Dagut & Laufer, 1985; Laufer & Eliasson, 1993; Darwin & Gray, 1999; Waibel, 2008; González, 2010; Ke, 2013; Gilquin, 2011, 2015b) has recurrently pointed out that MWVs are a "peculiarity of the Germanic languages" (Liao & Fukuya, 2004, p.211), a group to which Chinese does not belong.

At first sight, the higher MWV frequency in Chinese L2 production could be attributed to the presence of verb compounds in such a language (see section 2.4.2) that might behave semantically like literal MWVs in English (Chen, 2008; Biq, 2009; Zhang & Rao, 2012). However, as highlighted in previous research (Liao & Fukuya, 2004; Ke, 2013), the so-called Chinese particle verbs are neither structural nor semantic parallels to the English MWV. Hence, the argument that MWVs exist in Chinese does not help to explain the higher frequency of MWVs in the speech of speakers from this particular L1 background. Moreover, a group of verbal structures similar to MWVs has also been

extensively described for Italian (Massini, 2005; Iacobini & Masini, 2006; Iacobini, 2009a) (see section 2.4.3), yet data from the Italian speakers in the present study exhibited the lowest frequency of MWV occurrence.

The higher use of MWVs by Chinese speakers appears to be a much more complex issue that could be attributed to two factors. First, it might be the result of Chinese speakers' ability to produce more verbal forms in general as previous corpus-based research has shown (Chen, 2008; Chen, 2013). In fact, a search for the lexical verbs in each L1 background revealed that in the TLC that is indeed the case, i.e. Chinese speakers produced overall more lexical verbs per million words (99,076) than Italian (80,967) and Spanish (92,413) speakers. A second alternative explanation for the high MWV frequency in Chinese speech might be a typological one. In typological classifications, Chinese has been found to resemble English in the way it encodes verbal-directional motion events (Talmy, 2000; Slobin, 2004). Along other elements, a motion event includes an object that moves and a path specifying the trajectory or position that the object follows (Talmy, 1985). For example, in the sentence *Mary came in*, *Mary* is the object that moves, *came* is the motion and *in* represents the path. In English, the path is often realised in the form of a particle like *up*, *out*, *off*, or *in*. In Chinese, path information is encoded in a secondary verb that attaches to and is inseparable from the main verb and has a locative meaning such as *chu* (i.e. exit) in the example *tui-chu* (i.e. *push out*) (Chen, 2008; Zhang & Rao, 2012) (see section 2.4.2). It is possible, then, that the separation of the particle (i.e. the path) from the verb (i.e. the motion) in English makes the MWV form more salient and thus more readily learned by Chinese speakers given that they encounter both elements – motion and path – together in Chinese. As research on noticing and salience in L2 vocabulary use has

shown, differences that the L2 speakers are able to identify between their L1 and the L2 are likely to draw their attention (Ellis, 2006; Mackey, 2006; Gass & Mackey, 2006). We can then speculate that Chinese speakers might have been more likely to notice this gap between the form of English MWVs and the one of verbs in their own language, which may have facilitated the use of MWVs in conversation.

Regarding MWV use by speakers from the other L1 backgrounds analysed, a possible reason for the infrequent use of MWVs by Spanish speakers might be linked to a lack of awareness of the existence of MWV forms in Spanish given that the group of MWVs in this L1 consists of a rather limited number of items (see section 2.4.4). Spanish speakers then might find it harder to acquire MWVs and thus avoid their use in their spoken production. The low frequency of MWV occurrence in Italian, however, might be due to a different factor: the similarity between Italian and English MWVs. As was mentioned in section 2.4.3, Italian MWVs form a productive verb category in contrast to what authors like Waibel (2008) have argued. Not only do Italian MWVs resemble English ones with respect to their form, but they can also display different levels of compositionality including completely opaque MWVs (see section 2.4.3). As has been the case of other L2 speakers whose L1s contain MWVs that are rather similar to English ones (e.g. Dutch) (Hulstijn & Marchena, 1989), it is possible that Italian speakers tended not to use English MWVs that resembled familiar Italian ones, opting for different verbs instead.

### 5.3.4. L1 background and the range of MWVs

As with proficiency, the study operationalised the range of MWVs in the three L1 backgrounds (i.e. Chinese, Italian and Spanish) as i) the most frequent MWVs across

L1s, ii) the most common lexical verbs functioning in MWVs per L1 background, and iii) the most frequent MWVs produced specifically by each L1 group. The findings based on the previous first two aspects tend to resemble the patterns observed for L2 proficiency. For example, as it was also the case with the range of MWVs across proficiency levels, a very large proportion of the MWV occurrences in each L1 group is made up of a small set of MWVs and lexical verbs. The set of lexical verbs that appears in most of the MWV occurrences in all L1s (i.e. *go, come, grow, get*, and *take*) is identical to the one found in the results related to proficiency. This is not surprising given that, as was argued in section 5.2.2, all of those verbs except *grow* represent some of the most frequent lexical verbs "in their own right" (Biber et al., 1999, p. 412). They are also known for exhibiting a wide lexical patterning (Viberg, 1996; Lennon, 1996; Altenberg & Granger, 2001; Granger, Paquot & Rayson, 2006) and for having multiple meaning senses which in turn increases their chances to be used by L2 speakers.

Going back to the most frequent MWV forms across L1 backgrounds, the study revealed that more than half of Italian and Spanish speakers did not produce any MWVs, whereas one third of Chinese speakers' production lacked these verb forms. On the one hand, these results highlight once again the difficulty that MWVs represent for L2 speakers, not only of different proficiency levels but also from various L1 backgrounds (Paquot & Granger, 2012). On the other hand, the fact that MWVs were absent in the speech of a large number of Italian and Spanish speakers is supported by earlier studies on the acquisition and use of MWVs by speakers from both L1s (Azzaro, 1992; Masini, 2005; Waibel, 2008; González, 2010). Nonetheless, the argument provided in previous studies (i.e. that the absence of MWVs in the speakers' L1 is responsible for MWV underuse in the L2) does not seem to satisfactorily explain the

considerable low use of MWVs by Spanish and Italian speakers in the present study. MWV forms have been observed in Spanish (see section 2.4.4). Despite this fact, it has been claimed that Spanish speakers are less likely to use MWVs on the grounds that Spanish lacks this particular kind of verbs (González, 2010). As for Italian, it has been claimed that "the near absence of such a verb type" (Waibel, 2008, p.129) as well as the semantic constraints surrounding the use of MWVs in Italian often result in an overall lower frequency of MWV occurrence and possible avoidance of the MWV category in L2 English by these speakers. Alternative explanations for the infrequent use of MWVs by Spanish and Italian speakers are discussed in section 5.3.3.

With respect to the MWVs that occurred with high frequency in specific L1 groups, the study provided some more evidence of the relationship between particular MWVs and certain topics that the L2 speakers discussed. More specifically, the topics that Chinese, Italian and Spanish speakers were assigned in the conversation task appeared to activate a group of semantically related MWVs (e.g. the topic of health and fitness elicited MWVs like *work out* and *stand up*). Plausible reasons why topics in the conversation task elicited semantically related MWVs in the speech of speakers coming from such different L1 backgrounds are provided in section 5.3.2 as they have also been found to explain topic effects on MWV use at different levels of proficiency. In addition to the reasons described in section 5.3.2, it could be argued that since the analysis of learner variables in this study was restricted to L2 production in two tasks only (i.e. discussion and conversation) (the rationale for this decision is explained in section 3.7), the number of topics would be more limited than if all four tasks were considered. As a result of the more limited range of topics, the effect of some of the topics can become more prominent.

### 5.3.5. The case of *grow up*

The present study provided evidence of the relationship that exists between MWV frequency and polysemy and between lexical verb frequency and productivity. This means that the most frequent MWVs found (e.g. *go out, come back, go back, go on*) tended to be the most polysemous and that the most frequent lexical verbs functioning in MWVs often occurred with many different particles (e.g. *go, come, get, take*). Nonetheless, the MWV *grow up* and the lexical verb *grow*, which were among the most frequent MWVs and lexical verbs found in this study, were exceptions to the previous findings. As far as *grow up* was concerned, it was the second most frequent MWV in the TLC. *Grow up* alone accounted for 9% of the total MWV occurrences in the corpus. This is a rather high percentage given that, except for *go out* which made up 12% of the total MWV occurrences, each of the remaining most frequent MWVs accounted for 6% or fewer of the MWV occurrences. Moreover, *grow up* featured prominently in the production of speakers at all levels of proficiency and from the three L1 backgrounds. Despite its high frequency, *grow up* was not found to be a particularly polysemous MWV (i.e. it had only three dictionary senses). Regarding the lexical verb *grow*, it consistently appeared in the top three lexical verbs functioning in MWVs across proficiency levels and in the top 5 across L1 backgrounds. Contrary to other highly frequent lexical verbs (e.g. *come, get*), *grow* appeared as one of the least productive MWV-forming verbs as it only combined with one particle (i.e. *up*).

Upon a closer inspection of the contexts where *grow up* was used, its high frequency seems to be related to two interconnected factors: its particular flexibility to be used across topics and L2 speakers' tendency to relate a wide range of topics to their personal experiences as children or teenagers. Three different meaning senses were associated

with *grow up* in the corpus: i) to become an adult, ii) to behave or think as an adult should, and iii) to flourish financially. The first meaning was by far the most frequent. As Examples 49 to 54 below show, many topics naturally led to the use of *grow up* with the first meaning (e.g. coming-of-age ceremonies, early memories, bullying, young girls in the fashion industry, teaching minors, and differences between child behaviour in the past and nowadays). However, speakers kept using *grow up* even to talk about topics that may not be directly related to the topic of growing up (Examples 55 to 58). In those cases, speakers used *grow up* as a resource to talk about their own personal experiences as children or teenagers or the ones of close relatives.

49. C: It's er when a Jewish boy **grows up**

    E: yes

    C: and reads the Torah that is like the Bible… (Spanish, B1, discussion)

    *Topic: Coming-of-age ceremonies*

50. C: Maybe it's a good memory maybe it's a bad memory erm for me

    E: mm

    C: I want to change the p= the pala= the place when I where I **grew up**… (Chinese, B2, conversation)

    *Topic: Early memories*

51. C: They're five times more likely to have a serious criminal record when they **grow up** and they're six times more likely to be incarcerated by the age of twenty four… (Spanish, C1, presentation)

    *Topic: Bullying in Mexico*

52. C: In addition there are lots of girls working as models models since when they are very little

    E: mm

    C: and I think that will create them lots of problems in the future

    E: mm

    C: because they r= have **grown up** in an environment which is not suitable… (Spanish, C2, presentation)

    *Topic: Young girls in the fashion industry*

53. C: I am severe but I always try to let them understand that what I say it's for for them not for me not because I want to to but because I want that they **grow up** in a civil way… (Italian, C1, discussion)

    *Topic: Teaching minors with criminal convictions in a vocational school*

54. C: I teach children so I see them **growing up** but they are different from child-childhood and much different from my er the I think erm children nowaday care about themselves more than my age… (Chinese, B2, discussion)

    *Topic: Differences between child behaviour in the past and nowadays*

55. C: When I was er was a child I was a chocolate girl and my classmate call me meatball so **growing up** I started er **growing up** I started to I I started to lose weight… (Italian, B1, discussion)

    *Topic: Cooking healthy food*

56. C: I'm going to talk about sports I think

    E: Which sports in particular?

    C: er in general because I have tried a lot of sports since I was born I **growing up** I've changed er a lot of sports and they are very different… (Italian, B2, discussion)

    *Topic: Sports*

57. E: Are you going to do are you con= going to continue teaching?

    C: mm and doing tourist guide

    E: ah

    C: er in the future erm because my children are are erm **growing up** and so maybe in the future I will have more time I hope so and so I will er continue teaching in the morning maybe… (Italian, B2, discussion)

    *Topic: Teaching history*

58. C: My presentation is about the current situation of Venezuela the principle reason is that having **grown up** er here in Spain I was born in Venezuela

    E: mm

    C: I have family in Venezuela and I love Venezuela the same way as I love Spain… (Spanish, C1, presentation)

    *Topic: Social situation in Venezuela*

The fact that g*row up* encodes a meaning that is basic to the human experience, i.e. becoming an adult, possibly allowed speakers from all proficiency levels and L1

backgrounds to relate their own personal experiences to discuss virtually any topic raised in the examinations. Initially it was suspected that there could be a link between the frequent use of *grow up* and a possible effect of cognitive maturity, in other words, that the majority of instances of *grow up* were produced by the younger speakers in the corpus (ages between 8-15). However, an exploration of the distribution of *grow up* across the different age groups proved that that was not the case since middle (36-50 years old) and young (20-35 years old) adults produced more instances of *grow up* per million words (i.e. 159 and 146 respectively) than the younger, 8-15-year-old speakers (i.e. 138).

Speakers' reliance on *grow up* to discuss many topics often resulted in the erroneous use of such MWV. As illustrated in Examples 59 to 61, some speakers seemed to be unaware of the semantic contribution of *up* to the meaning of the whole MWV combination and used *grow up* as a synonym of *develop, increase* or *grow*. Still in other cases, they seemed to have over-extended the meaning of *grow up* and used it as a semantic equivalent of *raise* or *bring up* (Examples 62 to 64). It should be noted that instances of both types of misuse were found in the production of speakers from all proficiency levels and L1 backgrounds. In the case of Italian speakers, the transitive use of *grow up* is likely an L1 transfer issue as the Italian equivalent of *raise* in the phrase *raise a child* is *crescere* (e.g. *crescere un bambino*), which can translate as *grow* (Waibel, 2008). Since in Chinese and Spanish *raise* and *grow* are realised as two distinct verbs, we can only speculate that speakers from both languages transitivized *grow up* in compensation for their lack of a better way to express the meaning of *raise*. In all instances, however, the transitive use of *grow up* demonstrates speakers' unawareness of the syntactic restraints of this particular MWV.

59. C: The young people er decided to start work in this festival

    E: right

    C: and er so er this festival **grow grow up** it grew and grew grew and **grew grewed up**… (Italian, B1, discussion)

60. C: But you cannot explain why for instance er erm unemployment w-was **growing up** is **growing up** and why erm public school no erm er private school have financed more than the public ones… (Italian, B2, discussion)

61. C: I think every day change the use of the internet I remember it was er something for only a company

    E: uhu

    C: now it's called an intranet only to communicate in the company and

    E: mm

    C: in the it er suddenly er **grew up**… (Spanish, C1 conversation)

62. C: I think that er visit your grandparents

    E: mm mm

    C: very important

    E: mm

    C: because

    E: mm

    C: they can give you

    E: mm

C: er a lot o-of erm of er advice they can er give you a lot of love

E: mm

C: because they **grew up** your erm your father or your mother… (Italian, B2, conversation)

63. C: The girl's

    E: mm

    C: family erm wi= erm give give the girls food and to erm erm er **grow grow her up**… (Chinese, B2, conversation)

64. C: The women the w=

    E: yeah

    C: and the women were supposed to stay at home with the family just I mean **growing** the children **up**… (Italian, C1, conversation)

## 5.4. The effect of task type on L2 MWV use

### 5.4.1. MWV frequency in the presentation, discussion, interactive task, and conversation

The results related to the effect of task type on L2 MWV use will be discussed in terms of the frequency and range of MWVs, which were examined across four tasks in advanced speech: presentation, discussion, interactive task, and conversation. Regarding overall MWV frequency, the study showed that L2 speakers used more MWVs in the presentation than in the other three tasks. There was very little variation in the MWV frequency of occurrence found in the discussion and conversation tasks, and the interactive task elicited the lowest number of MWVs. These findings are

somewhat surprising given that the presentation was a monologic task with minimum interaction between candidates and examiners aside from some backchannelling cues provided by examiners. A higher MWV occurrence was expected in the more interactive tasks as MWVs have been reported to occur much more frequently in conversational registers (Biber et al., 1999; Darwin & Gray, 1999). The higher frequency of MWVs in the presentation may possibly be explained in terms of the planning involved, which may have allowed speakers to more carefully choose the language to be used.

As opposed to the dialogic tasks, the presentation is the only pre-planned task. Not only do the speakers pre-select the topic of their presentation, but they also prepare brief notes to aid them during their oral delivery. In addition, as part of their examination, candidates are specifically required to use a "wide range of idiomatic expressions and colloquialisms" (Trinity College London, 2018, p.51). Therefore, it is possible that advanced L2 speakers, aware of the idiomatic properties of MWVs, made sure they included them in their presentation in order to meet some of the language requirements of the task.

The study also revealed two important findings with respect to the frequency of PPVs (e.g. *get back to, come up with*). First, the speech of advanced L2 speakers in general contained the majority of the PPV occurrences found in the corpus. Second, PPVs featured more prominently in the presentation task. Concerning the first of these findings, it is important to note that PPVs are less frequent and semantically more complex than other types of MWVs and tend to be mainly used in idiomatic senses (Claridge, 2000; Biber et al. 1999; Rodríguez-Puente, 2019), which might make them

more difficult for lower level speakers to use. The higher PPV frequency in the production of advanced speakers may then be related to the fact that highly proficient speakers have more knowledge of the language and in general a wider lexical repertoire which allows them to use more semantically sophisticated vocabulary than less proficient speakers (Ellis, Simpson-Vlach & Maynard, 2008; Erman et al., 2016; Alexopoulou et al., 2017; Webb, 2018). Regarding the higher frequency of PPVs particularly in the presentation task, the most likely explanation is that planning afforded speakers the opportunity to productively use PPVs which the speakers were presumably already familiar with. As Elder & Iwashita (2005) argue, planning allows L2 speakers to retrieve and later produce specific lexical forms and complex linguistic items, which might otherwise not be used when no planning is possible.

### 5.4.2. Range of MWVs in the presentation, discussion, interactive task, and conversation

As was mentioned in the results of this study (section 4.4.2), the range of MWVs was understood as i) the variety and distribution of the top MWVs across tasks and ii) the specific set of top MWVs per task. Considering the first aspect and in line with corpus-based research on MWV use in L2 contexts (Waibel, 2008; Chen, 2013), the results showed that the distribution of MWVs produced by the whole group of advanced speakers follows a Zipfian profile, with a small set of MWVs making up almost two-thirds of the MWV occurrences. Moreover, the results indicated that there seems to be a relationship between the speaking tasks and the distribution of MWVs in C1-C2 speakers' speech, with the presentation being the task where all the top MWVs occurred at least once.

As far as the specific set of top MWVs per task is concerned, the presentation consistently displayed a higher frequency of MWVs used in transitions within the context of oral presentations. Those transitions acted as discourse organising devices, and the MWVs in them realised different functions associated with keeping the coherent structure of an oral presentation, for example, summarising (Example 65), signalling the introduction of a new topic or idea (Example 66), emphasising an idea (Example 67), and returning to a previous point (Example 68). The fact that the presentation required an "effective and identifiable structure" (Trinity College, 2018, p.41) is likely to have contributed to the use of MWVs in transitions that helped advanced speakers organise the content during their oral delivery. In fact, upon a closer analysis, several of the MWVs in the presentation were used at the beginning of the task to clarify its structure. Examples 69 to 73 illustrate this pattern.

65. C: The tangle of the red shoes represent the entire so the elegance the sophistication and the joy that every woman should be able to express without any kind of limits er to **sum up** I'd like to say that I hope that the femicide will stop… (Italian, C1)

66. C: We can use the animal organ while we are waiting for a real human one however er we are giving the animal solution and as far as I'm concerned is er unfair to the animal er now I'd like to **move on** to the drawbacks of xenotransplantation… (Spanish, C2)

67. C: Based on the results of that exam the needs of the school itself and the children they apply strate= er strategies to be worked on during the school year and at this time

    E: mm

    C: I would like to **point out** that at the present time only in the state of Chihuahua there are about sixty five working in this area… (Spanish, C1)

68. C: Well we we hope this will improve in the future er also industry well sorry er the government to **go back** to the government congress has passed stronger laws against piracy… (Spanish, C2)

69. C: I have divided my topic into three sections

    E: mm

    C: first of all I'm going to talk about the of bullfighting then I'll **go on** to talk about the controversy around bullfighting… (Spanish, C1)

70. C: My my topic is about senior high school students in China going abroad to study and the fir-first of all I'd like to explain the reasons and then I will **go on** talking about the advantages and disadvantages and last will I will express my own opinions… (Chinese, C1)

71. C: The first part of my topic is with pros of adoption in these kind of families and then I would like to **move on** to the potential drawbacks er some nuclear families… (Spanish, C1)

72. C: I have divided my topic into three sections first I'm going to talk about the arguments that people who are for legalisation have to defend the legalisation then I would like to **move on** and speak about er my arguments and my beliefs and feelings about legalisation of drugs

    E: mm

    C: er then I'll turn the page and talk about er the use of drugs for medical purposes and to **sum up** I'll end with a conclusion… (Spanish, C1)

73. C: I will start my presentation by **pointing out** some negative effects of er piracy especially on the economic side er then I will continue on to speak about the organisation of this erm er activity… (Spanish, C2)

In addition to showing that the presentation elicited MWVs functioning in transitions, the study also revealed that several of the MWVs in the presentation were also found among the most frequent MWVs in the conversation, discussion and interactive tasks. However, rather than functioning as part of transitions, the MWVs in those tasks were used to express a wide variety of meaning senses. For instance, the meanings expressed by *go on*, *go back* and *move on* in Examples 74 to 76 do not relate to the ones that the same MWVs conveyed when used as part of transitions in the presentation task. What is more, the same MWV form was often found to express different meanings in the production of an individual speaker (Examples 79 and 80). These findings are relevant for two reasons. First, they provide evidence of advanced speakers' awareness of the polysemous nature of frequent MWVs. Second, they are indicative of advanced speakers' ability to adjust their MWV use in order to meet task demands. In earlier studies on MWVs and task types (e.g. Hulstijn & Marchena, 1989; Laufer & Eliasson,

1993; Sjöholm, 1995; You, 1999; Liao & Fukuya, 2004; Barekat & Baniasady, 2014; Becker, 2014), the extent to which L2 speakers can adapt their use of MWVs across tasks has been rarely discussed. This is due to the fact that the tasks selected for such studies (e.g. memorisation, translation, recall, story re-tell) (see section 2.5.1) were highly controlled and provided speakers with narrow communication contexts. However, this study showed that in spoken production tasks high-proficiency L2 speakers seem to be sensitive to the contexts where MWVs are used and are able to select the appropriate meaning senses of a MWV in accordance with the nature and requirements of the tasks in which they engage. As further illustrated by the examples below, the speakers can appropriately adapt their use of MWVs across tasks (Examples 77 and 78; both of these come from one speaker) and even within the same task (Examples 79 and 80; these examples also come from one speaker).

74. C: We can't live without media just like er a fish can't live without water so we have to know what the things are going are **going on** around us… (Chinese, C1, discussion)

75. C: Now more than ever erm there are cities and metropolis everywhere erm and def-definitely in my opinion we'll have to **go back** to farming… (Spanish, C1, conversation)

76. C: It's like a guide you know because er now we have to choose a university I have to **move on** and er in the past I had to choose and I use I used to think that er my goal guide me around this new travel… (Italian, C1, discussion)

Speaker 1, Spanish, C2

77. It's not just a matter of doing I dunno ha= working on on electric cars no it's an holistic approach so because the way you organise a city has an impact in different aspects another practical example let 's **go back** to the detached house I'm a promotor and I built detached house in the extension of the city…

    (Presentation)

78. C: I remember was like that when I was a kid

    E: yeah yeah

    S: and I think er I think we need to **go back** to those neighbourhoods

    E: mm

    C: because they were more secure of course they were more sustainable

    (Discussion)

Speaker 2, Spanish, C1

79. They can become more accepted by society er now I would like to **move on** to the disadvantages...

    (Presentation)

80. Another disadvantage but it is not considered er as important as the other ones is that children can be left out at school because their parents are homosexual or are single but many children are left out at school for several

reasons and they **move on** quickly so it is not considered as important as the other ones…

(Presentation)

Differences in MWV use between the presentation and the other dialogic tasks were also evident in MWVs that were not part of transitions. One example that bears special attention is *find out*. This verb was the second most frequent MWV in the presentation task, and it was also very frequently used in the discussion and conversation tasks. In the presentation, speakers produced *find out* mainly i) to emphasise that the speaker conducted further research on the topic of the presentation (Examples 81 and 82), and ii) to report what an expert or research group found and to use that information as support for the points being developed (Examples 83 and 84). In the dialogic tasks, *find out* was predominantly used when speakers simply shared a piece of information that they learned (Examples 85 and 86). Although actual knowledge of the content was not assessed in the presentation, we can speculate that given the formal nature of this task, advanced speakers felt the need to include support from sources for the points that they were developing. The verb *find out* helped them accomplish such purpose in that particular task.

81. C: The conclusions were that there was no significant health change so several other independent studies also confirmed that in fact I could I **found** this information **out** first which I think is a reliable source… (Spanish, C1, presentation)

82. C: Mobile phones have made singletasking seem really boring and a cell phone just gives them er an really opportunities to multitask so after researching deeper into this matter I have **found out** four impacts… (Chinese, C1, presentation)

83. C: When a group of scientist **found out** the of radio waves coming from a region of space and they finally **found out** that these waves were coming from a supermassive black hole in the centre of our universe… (Italian, C1, presentation)

84. C: I also done some research about these psychologists and the University of Alabama and Birmingham erm they they **find out** that they **find out** that stu= s= s= s= children are at a more high risk when they cross er more high risk for death and injuries… (Chinese, C1, presentation)

85. C: I think it's more you don't trust them because er what shall you see and what the er television er shows about them it's not what really it is er and afterwards we always **find out** that there were other things behind what they did and erm erm and so you you lose your trust er in them… (Italian, C2, conversation)

86. C: Like I see a plane I wanna know how it works I can just type plane on Google and **find out**… (Italian, C1, conversation)

With respect to the top MWVs that occurred exclusively in one of the dialogic tasks, the study provided some evidence of a topic-effect phenomenon. When certain topics recurred in the corpus, specific MWVs associated with those topics recurred as well. Evidence of the influence of topics on MWV use mainly came from the conversation and interactive tasks (see section 4.4.2). The reasons for the effect of task topic on MWV use described in section 5.3.2 also apply to the patterns found in the dialogic tasks. First, some topics (e.g. healthy lifestyles, role models) may naturally lead to the use of particular MWV forms that typically express some of the concepts associated with such topics (e.g. *work out* and *look up to* respectively). Second, some MWVs may have been taught as topical vocabulary that the speakers could use to discuss topics that they later encountered in the interactive and conversation tasks (e.g. youth behaviour and social issues). In addition to those two reasons, the recurrent use of prompts or topics that already contained MWVs may also explain the frequent use of particular MWVs in the interactive and conversation tasks (Gablasova, Brezina & McEnery, forthcoming).

Cases of MWVs that the L2 speakers picked up from the topics or prompts that were used in the interactive phase and conversation deserve special attention. It is well known that speakers tend to repeat input from their interlocutors in conversation (Biria & Golestan, 2010; Carbary, Frohning & Tanenhaus, 2010). Thus, it was not altogether surprising that advanced speakers kept using MWVs that they previously encountered in prompts or that were explicitly used by examiners when introducing topics (see section 4.4.2). In those cases, the MWVs were clearly associated with specific subject areas (e.g. *put down* was used in relation to the topic of whether it is humane to put animals down). Another important point regarding these verbs is that their use arguably

implies understanding of the MWV form given that in order to meaningfully reproduce a MWV after hearing it, one has at least to be familiar with its meaning and context of use.

Regarding inter-speaker variation reflected in the speech of the most prolific speakers (i.e. those producing five or more MWVs per thousand words in their examinations), no single task contained the largest number of MWVs. However, overall the frequency of MWVs in the presentation and the discussion tasks remained relatively high. The differences in the MWV use by advanced speakers can partly be explained in terms of what Granger (2019) calls the "group versus individual perspective" (p. 241). This means that the general patterns of MWV use observed at the advanced group level might not apply to individual speakers. Even though the presentation was generally the task eliciting most MWV tokens for the majority of advanced L2 speakers in the TLC, some individual speakers produced more MWVs in one of the dialogic tasks. Regarding possible reasons why the discussion also contained large numbers of MWVs, topic familiarity may play a role. At advanced levels of proficiency, the discussion task is based on the ideas developed in the presentation task. Hence, it is likely that being already familiar with the topic to be discussed somehow facilitated the use of MWVs or even allowed speakers to reuse the MWVs that they had already employed in the presentation as opposed to the topics that were assigned in the other dialogic tasks.

### 5.5. Summary

This chapter examined the findings related to MWV use in L2 spoken communication in the TLC. It set out to compare PV and PPV frequencies with the ones reported in previous studies on MWV use across registers as well as L1 and L2 English varieties.

Next, the prevalence of a small group of MWVs, lexical verbs and particles accounting for most MWV tokens was explained in terms of the Zipfian distribution that characterises vocabulary use, suggesting that Zipf's law not only operates at the single word level as it has been previously reported but also at the phraseological multi-word verb level. Concerning non-canonical forms, the chapter addressed the four patterns of use identified for non-canonical MWVs. It then highlighted speakers' phraseological knowledge and creative potential to come up with unattested MWV forms. The analysis of the results related to polysemy also threw light on the extensive MWV sense knowledge that L2 speakers as a group possess given that they were found to be familiar with a wide range of both key and non-key meaning senses.

The chapter then moved on to discuss the effect of learner variables on MWV use. In terms of L2 proficiency, the results indicating that the main effect of this variable on MWV use was not statistically significant were found to contradict earlier findings regarding higher proficiency speakers producing more MWVs than lower proficiency ones. This difference in the results was taken as evidence of the complex relationship between L2 proficiency and MWV use, which was linked to the difficulty that MWVs pose to speakers at all stages of language development, the various definitions of the construct of L2 proficiency, and the level of individual variation present within a single proficiency group. With respect to the range of MWVs per proficiency level, the study revealed a high degree of overlap with speakers of all proficiency bands relying on a set of highly frequent MWVs. The overlap was attributed to the functionality of such MWVs and to the fact that they might represent 'safer' lexical choices to the speakers.

As far as L1 background is concerned, findings about the higher MWV frequency in the production of Chinese speakers did not align with previous claims on cross-linguistic differences and MWV use. Two possible reasons were provided to explain the high frequency of MWV occurrence in Chinese L2 speech, namely, Chinese speakers' ability to produce many verbal forms in general and the salience of MWVs. With respect to the other two L1 backgrounds, MWVs were found to occur infrequently in the production of Spanish and Italian speakers. In the case of Spanish speakers, it was argued that the low MWV frequency could be tied to Spanish speakers not being aware of the existence of MWVs in this L1, which could make their learning more difficult. Italian speakers, however, might have produced few MWVs because they considered them rather similar to MWV forms in Italian. The analysis of both L2 proficiency and L1 background also showed a strong effect of task topic with particular topics consistently eliciting specific MWVs.

Finally, the last sections of the chapter explored the significant effect of task type on MWV use and the possible reasons why the only monologic task contained most MWVs produced by advanced speakers. The use of MWVs to signal transitions between different sections of the presentation (e.g. *let's move on to*) seemed to have been encouraged by the requirements of the task. Moreover, the fact that advanced speakers varied their use of MWVs across tasks (i.e. MWVs served as transitions in the presentation but expressed a wider variety of meaning senses in the dialogic tasks) and even within the same task highlighted their ability to adapt their MWV use in accordance with task demands.

# Chapter 6: Conclusion

## 6.1. Introduction

This chapter first presents a summary of the main findings of this investigation (section 6.2). Following this summary, the theoretical and methodological contributions of the study are discussed in section 6.3 in accordance with the research aims outlined in section 1.5. Pedagogical implications following the main findings of the study are then outlined in section 6.4. The final sections of the chapter (sections 6.5 and 6.6) are devoted to a description of the limitations of this research and possible further research avenues on the topic of MWVs in L2 production.

## 6.2. Review of main findings

This thesis investigated L2 speakers' use of MWVs in spoken communication and the effects that L2 speaker-related factors (i.e. L2 proficiency and L1 background) and the type of speaking task have on this use. Three research questions guided this study. RQ1 aimed to describe the overall trends in the use of MWVs in the TLC. More specifically, RQ1 looked at four aspects related to the use of MWVs: i) MWV frequency and coverage, ii) lexical verb and particle productivity, iii) non-canonical MWVs, and iv) polysemy. With respect to the first of these aspects, the findings revealed that the frequency of PV occurrence in the TLC was relatively lower than those reported in previous corpus-based studies on MWVs in L2 spoken and written communication as well as in L1 spoken and written registers. The low PV frequency in the TLC was explained in terms of the nature of communication (i.e. the interaction in the TLC took place in a high-stakes context, and speakers lacked opportunities to plan their output)

and methodological factors (i.e. various ways to operationalise the term PV and differences in the approach that previous studies adopted when manually coding MWVs). As far as PPVs are concerned, their frequency of occurrence was much lower than that of PVs. This pattern was attributed to the overall low PPV frequency in English and the difficulty that PPVs might pose to L2 speakers because of their idiomatic and opaque meanings.

As regards MWV coverage, lexical verb and particle productivity, the findings showed that MWVs in the TLC follow a Zipfian distribution with a small set of MWVs accounting for more than half of the MWV occurrences. The same pattern was observed in the use of lexical verbs and particles (i.e. a small number of items made up the majority of MWVs). The most frequent lexical verbs that the study identified (i.e. *go, come, grow, get* and *take*) were also found to be the most productive ones (except for *grow*) given the wide variety of particles that each of them combined with. Regarding particles, *up* and *out* appeared in more than half of the MWVs and were particularly productive. The high frequency of both particles was attributed to the fact that they can express multiple meanings when combined with different lexical verbs as well as to their tendency to function as particles rather than as any other grammatical form.

The next aspect that the study looked at is that of non-canonical MWVs. The majority of non-standard MWVs occurred in the speech of the more proficient speakers (i.e. speakers from the B2 and C1-C2 groups), which was attributed to the better command of the L2 and wider linguistic repertoires of these speakers in comparison to lower-level L2 speakers. Four patterns of non-canonical MWV use were identified: i) MWVs containing a redundant particle (e.g. *restart over*), ii) MWVs with a lexical verb whose

pronunciation is close to one of other similar verbs in attested MWV forms (e.g. *sack up vs. stack up*), iii) MWVs with a particle contributing a special (often perfective) meaning (e.g. *please up*), and iv) combinations of verbs and particles that the speakers put together to express new meanings (e.g. *become out*). Even though some of the non-canonical MWVs that the speakers produced could be considered errors, others seemed to be examples of creative language use and made perfect sense in the context where they were used.

The last aspect that RQ1 explored was polysemy. In this regard, the findings indicated that the most polysemous MWVs in the TLC were among the most frequent ones occurring in the corpus. Although the range of meaning senses for the most polysemous MWVs was between four and eight, only one or two key senses dominated in their use. The predominant use of one or two key senses was linked to the fact that, regardless of how polysemous a MWV is, some meaning senses tend to be more salient and frequent than others. Furthermore, the findings also showed that the key senses of highly polysemous MWVs in the TLC appeared to broadly resemble those reported for the same verbs in L1 speech.

Despite findings showing a close relationship between MWV frequency and polysemy, *grow up* was one of the most frequent MWVs across all L2 proficiency levels and L1 backgrounds, yet it was not particularly polysemous. In the same vein, the lexical verb *grow* was among the most frequent lexical verbs; however, it was one of the least productive ones. The high frequency of *grow up* was attributed i) to the fact that it appeared to be widely usable across contexts and ii) to speakers' tendency to relate a wide range of topics to past experiences from their childhood and teenage years.

In response to RQ2, which sought to explore the effect of learner variables on L2 speakers' MWV use, the findings revealed that overall there was a relatively low MWV representation in L2 production across the proficiency levels and L1 backgrounds studied. On average, speakers from these proficiency levels and L1 backgrounds produced no more than one MWV per thousand words. L2 proficiency was not found to be a statistically significant factor in MWV use. It appeared that becoming more proficient in the L2 was not related to an increase in the frequency of MWVs, a finding that was not in line with what earlier research reported. This study argued that the relationship between L2 proficiency and MWV use is rather complex and that differences between previous findings and the ones obtained in this investigation appeared to be related to the difficulty that MWVs represent for speakers at all proficiency levels, the different definitions of L2 proficiency adopted in previous studies, and the relatively high level of inter-speaker variation in L2 MWV use that may be found within a single proficiency group.

The second learner variable that RQ2 examined was L1 background. Unlike L2 proficiency, the effect of L1 background on MWV production was found to be statistically significant. L2 speakers from a Chinese L1 background produced more MWVs per thousand words than Italian and Spanish L1 speakers. It was argued that Chinese speakers' ability to produce more verbal forms in general and the salience of MWVs in English might explain why speakers from this L1 background produced more MWVs than speakers from the other L1 groups.

Regarding the relationship between learner variables and the range of MWVs, the study brought to light the following main findings. First, the coverage patterns of MWVs and lexical verbs observed for MWV production in the TLC were also found to be true for MWV use at all proficiency levels and L1 backgrounds (i.e. a small set of MWVs and lexical verbs made up the majority of MWV occurrences). Second, a high degree of inter-speaker variability characterised MWV production by speakers from the same proficiency level and L1 group. Some speakers produced a large number of MWVs whereas others did not use any. Third, L2 speakers from all proficiency levels appeared to consistently rely on a set of highly frequent MWVs, which was linked to the functionality of these verbs and the fact that they might have represented safer lexical choices for B1, B2 and C1-C2 speakers. In contrast, a much lower degree of overlap was observed in the use of the most frequent MWVs by speakers from the three L1 backgrounds analysed. Fourth, the topics that L2 speakers discussed, particularly in the conversation task, appeared to activate the use of specific MWV forms, thus providing evidence of a topic-effect phenomenon in L2 MWV production.

Finally, RQ3 investigated the effect of task type on advanced L2 speakers' MWV use. Four speaking tasks were explored: presentation, discussion, interactive task, and conversation. In terms of MWV frequency, the monologic presentation contained most of the MWVs that advanced speakers produced, and it was also the task where L2 speakers used more MWVs per thousand words compared to the other dialogic tasks. As regards the range of MWVs, the most frequent MWVs in the presentation task tended to act as part of transitions that helped speakers organise the content of their presentations and realised different purposes (e.g. summarising, introducing a new topic or idea, emphasising an idea, and returning to a previous point). Many of these high

frequency MWVs in the presentation were also found to occur very frequently in the other three dialogic tasks (i.e. discussion, interactive task and conversation); however, they were used to express a wide variety of meaning senses rather than to serve as part of transitions. What is more, advanced speakers appeared to be able to use the same MWVs to express different meanings within the same task, which provided evidence of speakers' familiarity with different MWV meaning senses and their ability to adapt their MWV use. In the interactive task and the conversation, MWV production tended to be influenced by the recurrence of certain topics. The effects of task topic in these two dialogic tasks appeared to be even more evident in cases where the topics and prompts discussed already contained MWVs, which the L2 speakers kept using in their interactions.

## 6.3. Contributions of the study

The purpose of this research was to describe MWV use by L2 speakers in an interactive, spoken context and the effects of three variables on this use (i.e. L2 proficiency, L1 background, and type of speaking task). The study was based on data from a large corpus representing L2 speech and adopted both qualitative and quantitative approaches to analyse this data. The contributions that the study has made can be divided into two types: theoretical and methodological.

### 6.3.1. Theoretical contributions

The theoretical contributions of the study relate to three specific areas, namely, the role of i) register, ii) learner variables, and iii) task effects. Concerning the first of these areas, one of the main contributions of this study has been advancing our understanding

of the patterns of L2 MWV use in naturally-occurring, spoken communication, a register that has been little explored in previous elicitation and corpus-based studies on MWVs. In contrast to the relatively high MWV frequency reported for L2 writing (e.g. Waibel, 2008; Deshors, 2016), the findings indicated that overall there was a low MWV representation in spoken L2 production. Despite this relatively low MWV frequency, the distribution of MWVs in spoken L2 production appeared to resemble the one reported for L1 spoken and written registers (e.g. Gardner & Davies, 2007; Liu, 2011), with a few MWV forms making up the majority of MWV occurrences.

Another theoretical contribution of the study relates to the insights that it has provided into L2 speakers' MWV knowledge by showing that they were able to create meaningful MWVs (see description of non-canonical MWVs categorised as patterns three and four, section 4.2.3). As explained in section 2.3.2.4, novel MWV forms in L2 contexts have been studied mainly using written corpora such as the ICLE (e.g. Gilquin, 2015a; Schneider & Gilquin, 2016); however, this study has shown that L2 speakers used their knowledge of lexical verbs and particles to create MWVs when participating in spontaneous, spoken interactions.

With respect to the role of learner variables in MWV use, this study focused on two particular variables: L2 proficiency and L1 background. As was already mentioned in section 5.3.1, the evidence establishing a link between L2 proficiency and MWV production has mainly come from elicitation studies in which the use of selected MWV forms by high-proficiency speakers was assessed (e.g. Liao & Fukuya, 2004). Therefore, the main contribution of the present study in terms of the influence of L2 proficiency on MWV production has been providing a description of how the frequency

and range of MWVs vary across different proficiency levels, including the production of lower-level L2 speakers. Moreover, the study has helped to understand that the relationship between L2 proficiency and MWV use tends to be complex, rather than linear, and highlighted the major role that inter-speaker variation within the same L2 proficiency group plays in MWV use. Regarding the L1 background variable, previous research pointed out that L2 speakers whose L1 contains MWVs or similar forms tend to produce more MWVs than speakers whose L1 lacks this type of verb (e.g. Siyanova & Schmitt, 2007; González, 2010). This study, however, revealed that the mere presence of the target structure in speakers' L1 (i.e. MWVs were found in Chinese, Italian and Spanish) may not be sufficient for L2 speakers to productively use MWVs in spoken communication. This finding further contributes to the current debates about the extent to which L1 background has a prominent role in SLA (e.g. VanPatten & Williams, 2015; Ringbom, 2016).

The third area concerns task effects on MWV use. As was pointed out in section 1.6, at the time of writing this thesis there had not been any attempt to explore the effect of task type on L2 MWV production using a corpus-based, rather than a corpus-informed, approach. This study demonstrated that different speaking tasks shape MWV production in several ways. First, the findings showed that the presentation, a task that allowed for planned output and can be considered typical of academic contexts, contained the highest frequency of MWVs, including the largest number of PPV forms. Second, specific MWVs were found to consistently feature in tasks based on discussions of particular topics (e.g. *hang out* recurred in the context of youth behaviour), a finding that highlights the importance of task topic effects on L2 vocabulary use (Alexopoulou et al., 2017; Gablasova, Brezina & McEnery,

forthcoming). Finally, the study revealed that advanced speakers are not only aware of the multiple meaning senses of frequent MWVs but are also able to select the appropriate meaning sense of a MWV in accordance with the context of speaking and type of tasks in which they engage.

### 6.3.2. Methodological contributions

This study analysed MWVs produced by L2 speakers that differed in terms of their level of L2 proficiency (B1, B2, and C1-C2) and L1 background (Chinese, Italian, and Spanish). While most earlier learner corpus-based work on MWVs has been based on comparisons between L1 and L2 speakers' MWV production (e.g. Chen, 2013; Gilquin, 2015b), this study adopted an L2-L2 comparative approach (see section 1.3.2) (Granger, 1996, 2015). This comparison of interlanguage varieties allowed the observation of MWV use patterns that are characteristic of the L2 groups studied.

One crucial contribution of this study relates to the area of data analysis, more specifically to the approach followed to explore the nature of L1 background influence on MWVs. First, in response to calls for a more rigorous study of L1 background (Jarvis, 2000; Paquot, 2013), the present study acknowledged the presence of MWVs and similar verbal forms in Chinese, Italian and Spanish, the latter being commonly reported as an L1 that lacks MWVs (e.g. González, 2010). Second, the study also provided a description of the form as well as the syntactic and semantic properties of MWVs in the three L1 backgrounds analysed. This description proved revealing because it made it possible to observe i) where the differences between the MWV forms in speakers' L1 and in the L2 lie, and ii) the extent to which these differences accounted for MWV use by Chinese, Italian and Spanish speakers.

A central issue in the methodology of this study has been making the processes of data extraction and manual treatment as transparent and replicable as possible. In this respect, another important contribution of this study has been offering a framework for the manual coding of MWVs, which constitutes an essential step in the analysis of MWVs in L2 contexts. As was explained in section 3.10.2, the automatic extraction of MWVs presents a challenge to researchers due to the fact that often non-MWV structures make their way into the query results. This study has presented a comprehensive description of the coding scheme and the process of manually filtering the results using this scheme. Moreover, the manual analysis was followed by a double-coding process and a calculation of the inter-rater agreement level, which proved an effective way to confirm the reliability of the coding. It is hoped that the framework for the manual analysis of MWVs developed in this study can be used and adapted by future studies on L2 MWV production.

### 6.4. Pedagogical implications

This study primarily focused on describing MWV use by L2 speakers in spoken, interactive communication. Although its aim was not exploring the pedagogical dimension of MWV use (i.e. it did not look at a particular approach for the teaching of MWVs to L2 speakers), several pedagogical implications derive from the findings of this study. First, this investigation has shown that in naturally-occurring communication the frequency of MWVs in L2 speech tends to be relatively low. The immediate implication that this finding has for the teaching of MWVs is that more attention could be directed towards helping L2 speakers at all levels of L2 proficiency to confidently use more MWVs when engaging in spoken interactions.

Next, with respect to the range of MWVs, the findings indicated that when it comes to the most frequent MWVs, it is relatively the same set that is consistently used by B1, B2 and C1-C2 speakers. Therefore, given the multiple benefits that MWVs can bring to L2 speakers (see section 1.2), it seems that an intervention is desirable to help them progressively expand their repertoires of MWVs as they become more proficient in the L2. This type of pedagogical intervention can also focus on contextually appropriate MWV use as having a large number of MWVs in L2 production does not necessarily mean that one can use MWVs in semantically and pragmatically appropriate ways (Waibel, 2008). Moreover, when helping speakers from different proficiency levels expand their range of MWVs, it is important to consider inter-speaker variation. As was shown in this study, a large inter-speaker variation can be found even within one proficiency level, which may result in a rather heterogenous MWV knowledge base.

It has been pointed out that the lack of lexical correspondence between MWVs and verbs in speakers' L1 is one of the main sources of difficulty to learn MWVs and can lead to MWV avoidance. As González (2010) puts it "cross-linguistic influence may be posited as one of the driving forces behind the avoidance of this group of verbs" (p.153). This study reported that MWV-like structures exist in Chinese and in two Romance languages (i.e. Italian and Spanish). In fact, MWVs in Italian were found to be part of a productive verb category and closely resemble English MWVs in terms of their form and semantics (see section 2.4.3). Language practitioners can then take advantage of MWV descriptions such as the ones provided in this study and develop L2 speakers' awareness of the existence of MWVs in their native language, which might facilitate speakers' learning of a verb type that is often perceived to be exclusively English

(Thim,2012).

Other pedagogical implications follow from the findings related to topic effects and the level of task planning. The effect of topic on MWV production was observed across proficiency levels and L1 backgrounds. Specific MWV forms were repeatedly used in connection with particular subject areas. It is then advisable to consider selecting various topics and prompts for discussion in L2 classrooms so that L2 speakers have greater opportunities to use different MWVs. On a similar note, the fact that the only pre-planned task in this study contained the highest MWV frequency including the majority of PPVs, which are considered to be semantically complex (Claridge, 2000), has further pedagogical implications for the selection and design of tasks used in both testing situations and classroom settings. For testing purposes, it seems that the level of planning that a task requires should be carefully considered if L2 speakers are expected to demonstrate knowledge of a wide variety of MWVs. For teaching purposes, more attention should perhaps be given in the L2 classroom to help L2 speakers expand their repertoires of MWVs when performing in unplanned, spontaneous output-based tasks.

Finally, language practitioners and teachers often struggle to select an appropriate approach to the teaching of MWVs to L2 learners (Schmitt & Redwood, 2011). Proposals to teach MWVs to L2 users abound in the literature (e.g. Side, 1990; Darwin & Gray, 1999; Dirven, 2001; Kurtyka, 2001; Armstrong, 2004; Condon, 2008; Yasuda, 2010; White, 2012; Birjandi, Alavi & Najafi, 2015; Torres-Martínez, 2015; Ke, 2017). While having such a wide range of options is desirable given that no single approach is appropriate for all teaching contexts, it might make the selection more difficult. Prior to deciding which approach to use, it is important that teachers first understand the

extent to which individual factors such as L2 proficiency and L1 background may affect their students' learning of MWVs. The findings of the present study can contribute to this understanding and be used to inform the selection of the MWV teaching approach that best meets the needs of L2 learners.

### 6.5. Limitations

This study was successful in providing a detailed description of MWV use by L2 speakers as well as the effect of learner variables and task type on MWV use. While all research aims were achieved, the following limitations can be pointed out. First, although several measures were taken to retrieve all possible MWVs in the TLC (see section 3.10.2), L2 speakers still appeared to produce relatively few MWVs. Therefore, the study could have benefited from a larger size of the evidence (e.g. longer individual contributions) in order to explore, for example, whether there are larger differences between the MWV frequency in B1, B2 and C1-C2 speakers' production.

Second, since this study focused on MWV use across speaking tasks and by L2 speakers from different proficiency levels and L1 backgrounds, L2-L2 comparisons (i.e. comparisons between two or more interlanguage varieties) proved to be the most suitable methodological approach. MWV production by L2 speakers was not compared to the one by L1 speakers. For a meaningful comparison between L1 and L2 speakers' language production, one essential requirement is that the data come from comparable corpora (Granger, 2015; Gablasova, Brezina & McEnery, 2017). Considering the discourse type, register and task structure of the TLC, a fully comparable L1 spoken corpus, which would make the comparison meaningful, was not available. Even though an L1 spoken corpus that can be directly compared to the TLC is currently being

developed (i.e. the Trinity Lancaster Corpus for L1 English Interaction, TLC-L1) (Gablasova, Brezina, McEnery, forthcoming), this resource was not available at the time of writing this thesis.

Third, an analysis of the different semantic categories of MWVs would be useful in understanding MWV compositionality in L2 production and could shed light on the relationship between learner variables such as L2 proficiency and the different levels of semantic opacity (e.g. whether gains in L2 proficiency go hand in hand with a more frequent use of non-compositional MWVs). However, this type of analysis was not carried out in the present study.

Lastly, the study could have benefited from an even larger range of proficiency levels including lower levels than B1.2 which could have allowed the observation of emergent patterns of MWV use in L2 speech. In addition, given the data available, the most advanced C1 and C2 speakers had to be aggregated (see section 3.8). It is possible, however, that if there had been sufficient numbers of speakers from both proficiency bands, different patterns of MWV use could have been observed. Most speakers from the C1-C2 group came from the lower C level (i.e. C1.1); therefore, it would be especially valuable to observe MWV production in the highest proficiency level (i.e. C2) to explore whether there is an increase in the frequency and range of MWVs used by the most proficient speakers in the corpus.

### 6.6. Further research avenues

The findings of this study can lead to further research in four main areas: i) L1-L2 comparisons of MWV production, ii) inter-speaker variation, iii) non-canonical

MWVs, and iv) the role of L1 background in MWV studies. First, future studies can adopt an L1-L2 comparison approach to determine the extent to which the patterns of MWV use observed in this study resemble those found in L1 production. As mentioned in section 6.3, a reference L1 corpus for the TLC is currently under development (Gablasova, Brezina & McEnery, forthcoming). Once it is publicly available, this new corpus can be used to look at, for example, the effect of task type on the frequency and range of MWVs. Two of the main findings of this study were that the only monologic, pre-planned task elicited most the MWVs that L2 speakers produced and that these MWVs tended to function as transitions in the context of delivering oral presentations. Therefore, future studies could further explore MWV distribution across the different monologic and dialogic task types and assess whether L1 speakers also tend to use MWVs to signal transitions in the presentation task. Moreover, a comparison with L1 data could also serve to explore whether the topic-effect phenomenon observed in L2 speakers' production of MWVs is also present in L1 speech (i.e. if specific MWV forms tend to be consistently used in association with certain topics). What is more, this exploration could be taken one step further in order to observe if L1 speakers produce the same MWV forms as L2 speakers in response to similar topics. On a similar note, a direct L1-L2 comparison of frequency patterns could also prove useful in determining the typical rate of MWV use in all four tasks and shed light on whether L1 MWV production also exhibits large inter-speaker variation.

Next, this study provided evidence of the high degree of inter-speaker variation that can be found in MWV production by speakers from the same L2 proficiency level. For example, while some B2 speakers produced many MWVs, others from the same proficiency group did not use any MWV in their interactions. Inter-speaker variability

is an important area that is worth pursuing in the future. Studies could profit from obtaining finer-grained distinctions of L2 proficiency to evaluate MWV use. In the TLC, this could be done by examining L2 speakers' overall achievement marks (i.e. distinction, merit, and pass) in each proficiency band. The results of this type of investigation could then contribute to determining whether there is any tendency for L2 speakers who obtained higher achievement marks (e.g. B2 speakers whose performance was awarded a distinction mark) to produce more MWVs than L2 speakers from the same proficiency group whose overall performance was given a lower mark (e.g. B2 speakers whose performance was awarded a pass mark).

Another research area that can be further pursued is that of non-canonical MWV use. This study found a total of twenty non-canonical MWVs, most of which corresponded to one-off occurrences. Given the low frequency of these non-canonical MWVs, it was not possible to observe if they were systematically used by different speakers. It would be important for future research to determine if these particular non-canonical MWVs could be considered isolated errors produced by one speaker or instances of lexical innovations that are conventionalised to some degree and are used on a systematic basis. Studies could confirm the presence of the non-canonical MWVs that this study identified in other L2 spoken corpora (e.g. LINDSEI) or even in large L1 corpora (e.g. TLC-L1, Spoken BNC2014) (Love et al., 2017). The occurrence of the non-canonical MWVs in other data sets could then be used to argue for their conventionalisation and systematic production. Along the same line, future studies can also dig deeper into non-canonical MWV use across 'outer circle' and 'expanding circle' varieties of English (Kachru, 1985). For instance, the four patterns of non-canonical MWV use found in this study (see section 4.2.3) could be compared to the ones observed in MWV

production by other EFL (e.g. French and Japanese) and ESL (e.g. Indian) speakers in the TLC.

Finally, this study found a connection between L1 background and MWV use. Chinese speakers appeared to use more MWVs than Italian and Spanish speakers despite the fact that MWVs can be found in the three L1s. The role of L1 background in the context of MWV production by L2 speakers could then be explored further by looking at the relationship between L1 background and MWV compositionality. An analysis can be performed to determine if speakers whose L1 has mostly compositional MWVs (e.g. Chinese and Spanish) in which particles mainly express direction or location (e.g. *go back*, *come in*) would also tend to produce compositional MWVs in the L2. This investigation would then contribute to a more holistic understanding of the role of L1 background in the production of MWVs.

### 6.7. Concluding remarks

This study was descriptive in nature. Its aim was to contribute to the line of research that combines SLA theories and learner corpus linguistics in order to explore both qualitatively and quantitatively the MWV knowledge that L2 speakers possess. In doing so, the study demonstrated that patterns in L2 MWV use result from a complex interplay between learner and situational variables. It is hoped that the work in this thesis can inspire further corpus-based work on the topic of MWVs in spoken L2 production, which represents a fertile research area.

**Appendix**

Absolute frequencies of MWVs found in the TLC (organised in descending order)

| | |
|---|---|
| go out | 280 |
| grow up | 207 |
| come back | 153 |
| go back | 88 |
| go on | 87 |
| find out | 63 |
| wake up | 61 |
| come out | 55 |
| get up | 54 |
| give up | 40 |
| take out | 36 |
| move on | 32 |
| come on | 31 |
| pick up | 31 |
| keep on | 29 |
| sum up | 29 |
| take off | 28 |
| bring up | 27 |
| go up | 26 |
| make up | 26 |
| dress up | 25 |
| go down | 25 |
| sit down | 23 |
| turn off | 22 |
| end up | 21 |
| fall down | 21 |
| carry on | 19 |
| carry out | 19 |
| set up | 19 |
| stand up | 18 |
| look up | 14 |
| get out | 13 |
| work out | 13 |
| come in | 12 |
| hang out | 12 |
| point out | 12 |
| build up | 11 |
| break up | 11 |

| | |
|---|---|
| come up | 11 |
| get on | 11 |
| set off | 11 |
| show off | 11 |
| stress out | 11 |
| come down | 10 |
| give out | 10 |
| help out | 10 |
| bring back | 9 |
| put down | 9 |
| take up | 9 |
| throw up | 9 |
| get back | 8 |
| give back | 8 |
| throw out | 8 |
| write down | 8 |
| come up with | 7 |
| figure out | 7 |
| get along | 7 |
| settle down | 7 |
| take back | 7 |
| turn out | 7 |
| break down | 6 |
| calm down | 6 |
| get off | 6 |
| stay up | 6 |
| take on | 6 |
| come out with | 5 |
| get in | 5 |
| go around | 5 |
| go in | 5 |
| leave out | 5 |
| look back | 5 |
| put on | 5 |
| put up | 5 |
| rise up | 5 |
| send off | 5 |
| shut down | 5 |

| | |
|---|---|
| switch off | 5 |
| turn on | 5 |
| act out | 4 |
| bring out | 4 |
| catch up | 4 |
| check out | 4 |
| cut down | 4 |
| cut off | 4 |
| fall over | 4 |
| go by | 4 |
| go on with | 4 |
| log in | 4 |
| look around | 4 |
| mix up | 4 |
| show up | 4 |
| sort out | 4 |
| speed up | 4 |
| take down | 4 |
| take over | 4 |
| turn around | 4 |
| bring about | 3 |
| bring down | 3 |
| burn up | 3 |
| fit in | 3 |
| go off | 3 |
| lay out | 3 |
| log on | 3 |
| mess up | 3 |
| miss out | 3 |
| open up | 3 |
| pay back | 3 |
| send back | 3 |
| send out | 3 |
| split up | 3 |
| start off | 3 |
| start out | 3 |
| stay in | 3 |
| stay out | 3 |
| tell off | 3 |
| think over | 3 |
| turn back | 3 |
| use up | 3 |

| | |
|---|---|
| warm up | 3 |
| wipe out | 3 |
| ask around | 2 |
| blow up | 2 |
| boil down to | 2 |
| bring in | 2 |
| brush up | 2 |
| carry over | 2 |
| check over | 2 |
| clean up | 2 |
| clear out | 2 |
| close down | 2 |
| come over | 2 |
| cover up | 2 |
| cut out | 2 |
| do up | 2 |
| fill in | 2 |
| fill up | 2 |
| get back to | 2 |
| go along with | 2 |
| go round | 2 |
| hand out | 2 |
| hang around | 2 |
| hold on | 2 |
| keep out | 2 |
| keep up | 2 |
| lay off | 2 |
| look down | 2 |
| look out for | 2 |
| look up to | 2 |
| make out | 2 |
| pass away | 2 |
| pour out | 2 |
| put off | 2 |
| rip off | 2 |
| save up | 2 |
| sell off | 2 |
| sell out | 2 |
| send in | 2 |
| set out | 2 |
| shout out | 2 |
| shut up | 2 |

| | |
|---|---|
| sit around | 2 |
| slow down | 2 |
| speak up | 2 |
| stand out | 2 |
| start in | 2 |
| stay on | 2 |
| turn down | 2 |
| turn up | 2 |
| answer back | 1 |
| back up | 1 |
| beat down | 1 |
| bind up with | 1 |
| branch out | 1 |
| break in | 1 |
| bring over | 1 |
| burn down | 1 |
| burst out | 1 |
| check off | 1 |
| check up | 1 |
| chop down | 1 |
| clamp down on | 1 |
| come along | 1 |
| conjure up | 1 |
| cool down | 1 |
| crack down | 1 |
| curl up | 1 |
| cut back | 1 |
| cut up | 1 |
| dance along | 1 |
| dig up | 1 |
| divide up | 1 |
| drift off | 1 |
| drop off | 1 |
| drop out | 1 |
| face up | 1 |
| fill out | 1 |
| flush out | 1 |
| get around | 1 |
| get by | 1 |
| get down | 1 |
| get through | 1 |
| get through to | 1 |

| | |
|---|---|
| get up to | 1 |
| go along | 1 |
| go in for | 1 |
| go through | 1 |
| hang up | 1 |
| hold back | 1 |
| hold on to | 1 |
| join in | 1 |
| jump up | 1 |
| keep down | 1 |
| keep up with | 1 |
| kick off | 1 |
| knock over | 1 |
| lay down | 1 |
| lead back to | 1 |
| leave in | 1 |
| leave over | 1 |
| let down | 1 |
| let off | 1 |
| live out | 1 |
| look down upon | 1 |
| look forward to | 1 |
| look out | 1 |
| make up for | 1 |
| meet up | 1 |
| mess around | 1 |
| move along | 1 |
| pass by | 1 |
| pass on | 1 |
| pass out | 1 |
| pick on | 1 |
| pick out | 1 |
| piss off | 1 |
| play around | 1 |
| plug up | 1 |
| print out | 1 |
| pull down | 1 |
| pull on | 1 |
| pull out | 1 |
| pull over | 1 |
| pull up | 1 |
| push around | 1 |

| | |
|---|---|
| push out | 1 |
| push over | 1 |
| put down to | 1 |
| put in | 1 |
| put out | 1 |
| put up with | 1 |
| reach out | 1 |
| rent out | 1 |
| roam around | 1 |
| rub off on | 1 |
| run out | 1 |
| run over | 1 |
| rush out | 1 |
| separate out | 1 |
| serve up | 1 |
| shake off | 1 |
| shake up | 1 |
| shoot out | 1 |
| sign out | 1 |
| sit up | 1 |
| sneak out | 1 |
| speak out | 1 |
| spill out | 1 |
| stand down | 1 |
| start over | 1 |

| | |
|---|---|
| start up | 1 |
| step up | 1 |
| stick on | 1 |
| stop over | 1 |
| suck in | 1 |
| suit up | 1 |
| switch on | 1 |
| take away | 1 |
| take in | 1 |
| tap in | 1 |
| think up | 1 |
| throw away | 1 |
| tidy up | 1 |
| tie back | 1 |
| trace back | 1 |
| track down | 1 |
| try on | 1 |
| try out | 1 |
| type in | 1 |
| walk in | 1 |
| wash out | 1 |
| watch out | 1 |
| wipe off | 1 |
| work up | 1 |
| wrap up | 1 |

## References


Alexander, R. J. (1987). Problems in understanding and teaching idiomaticity in English. *Anglistik und Englischunterricht, 32*, 105-122.

Akbari, O. (2009). *A corpus-based study on Malaysian ESL learners' use of phrasal verbs in narrative compositions* (Unpublished doctoral dissertation). Universiti Putra Malaysia, Serdang.

Akimoto, M. (1999). Chapter 7: Collocations and idioms in Later Modern English. In L. J. Brinton & M. Akimoto (Eds.), *Collocational and idiomatic aspects of composite predicates in the history of English* (pp. 207-238). Amsterdam: John Benjamins.

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning, 67*(S1), 180-208.

Alonso Alonso, R., Cadierno, T., & Jarvis, S. (2016). Crosslinguistic influence in the acquisition of spatial prepositions in English as a foreign language. In Alonso Alonso, R. (Ed.), *Crosslinguistic influence in second language acquisition* (pp. 93-120). Bristol: Multilingual Matters.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and application* (pp. 101-122). Oxford: Oxford University Press.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied linguistics, 22(*2), 173-195.

Anderson, S. R. (1992). *A-morphous Morphology*. Cambridge: Cambridge University Press.

Armstrong, K. (2004). Sexing up the dossier: A semantic analysis of phrasal verbs for language teachers. *Language Awareness, 13*(4), 213-224.

Ayadi, A. (2010). *Lexical translation problems: The problem of translating phrasal verbs. The case of third year LMD learners of English* (Unpublished master's thesis). Mentouri University, Algeria.

Azzaro, G. (1992). The syntactic learning of English phrasal verbs: Theory. *Rassegna Italiana di Linguistica Applicata, 24*(1), 33-60.

Bacchielli, R. (1999). An essential bibliographical guide to the synchronic and diachronic study of verb-particle combinations ('phrasal verbs') in English. *SLIN Newsletter, 20*, 4-21.

Baldwin, T. (2005). Deep lexical acquisition of verb–particle constructions. *Computer Speech & Language, 19*(4), 398-414.

Baldwin, T., & Villavicencio, A. (2002). Extracting the unextractable: A case study on verb-particles. In *Proceedings of the 6th conference on Natural language learning-Volume 20* (pp. 1-7). Association for Computational Linguistics.

Bannard, C. (2005). Learning about the meaning of verb-particle constructions from corpora. *Computer Speech and Language, 19*(4), 467-478.

Bardovi-Harlig, K. (2012). Formulas, Routines, and Conventional Expressions in Pragmatics Research. *Annual Review of Applied Linguistics 32,* 206-227.

Bardovi-Harlig, K. (2013). Developing L2 Pragmatics. *Language Learning, 63*(S1), 68-86.

Bardovi-Harlig, K. (2019). Formulaic language in second language pragmatics research. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.),

*Understanding formulaic language: A second language acquisition perspective* (pp. 97-114). Oxon: Routledge.

Barekat, B. & Baniasady, B. (2014). The impact of phrasal verb avoidance on the writing ability of the university EFL learners. *Procedia - Social and Behavioral Sciences, 98*, 343-352.

Becker, T. P. (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*(24), 1-33.

Benincà, P. & Poletto, C. (2006). Phrasal verbs in Venetian and regional Italian. In F. Hinskens (Ed.), *Language variation – European perspectives* (pp. 9-22). Amsterdam: John Benjamins Publishing Company.

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

Biq, Y. O. (2009). Locative particles in spoken Taiwan Mandarin. In J. Z. Xing (Ed.), *Studies of Chinese Linguistics: Functional Approaches* (pp. 133-153). Hong Kong: Hong Kong University Press.

Biria, R., & Ameri-Golestan, A. (2010). The impact of syntactic priming on English language learners' production: A transfer study of indirect questions/requests. *Journal of Language Teaching & Research, 1*(5), 583-590.

Birjandi, P., Alavi, S. M., & Najafi Karimi, S. (2015). Effects of unenhanced, enhanced, and elaborated input on learning English phrasal verbs. *International journal of research studies in language learning, 4*(1), 43-59.

Boers, F. (2013). Cognitive Linguistic approaches to second language vocabulary: Assessment and integration. *Language Teaching, 46*, 208–224 .

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245-261.

Bolinger, D. (1971). *The phrasal verb in English*. Cambridge, MA: Harvard University Press.

Booij, G. (2002). Separable complex verbs in Dutch: A case of periphrastic word formation. In N. Dehé, R. Jackendoff, A. McIntyre, & Urban, S. (Eds.), *Verb-particle explorations* (pp. 21-41). Berlin: Mouton de Gruyter.

Booij, G. (2007). *The grammar of words: An introduction to linguistic morphology*. 2nd edn. Oxford: Oxford University Press.

Booij, G. & van Kemenade, A. (2003). Preverbs: An introduction. In Booij, G. & van Marle, J. (Eds), *Yearbook of Morphology* 2003 (pp. 1-11). Dordrecht: Kluwer.

Brezina, V. (2018a). Chapter 12: Statistical choices in corpus-based discourse analysis. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp 259-280). Oxon: Routledge.

Brezina, V. (2018b). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.

Brinton, L. & Akimoto, M. (1999). Introduction. In Brinton, L. & Akimoto, M. (Eds), *Collocational and Idiomatic Aspects of Composite Predicates in the History of English* (pp. 1-19). Amsterdam: John Benjamins Publishing Company.

British National Corpus Consortium. (2000). British national corpus: World edition. Oxford: Humanities Computing Unit of Oxford University.

Brown, D. W. & Palmer, C. C. (2015). The phrasal verb in American English: Using corpora to track down historical trends in particle distribution, register variation, and noun collocations. In M. Adams, L. J. Brinton, & R. D. Fulk (Eds.), *Studies*

*in the History of the English Language VI: Evidence and Method in Histories of English* (pp. 71-97). Berlin: Mouton de Gruyter.

Caines, A. & Buttery, P. (2018). The effect of task and topic on opportunity of use in learner corpora. In V. Brezina & L. Flowerdew (Eds.), *Learner corpus research: New perspectives and applications* (pp. 5-27). London: Bloomsbury.

Callies, M. (2015). Learner corpus methodology. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp.35-56). Cambridge: Cambridge University Press.

Capelle, B. (2007). When 'wee wretched words' wield weight: The impact of verbal particles on transitivity. In M. Nenonen & S. Niemi (Eds.), *Collocations and idioms 1: Papers from the first Nordic conference on syntactic freezes,* Joensuu, Finland, 19–20 May 2006 (pp. 41-54). Joensuu: University of Joensuu.

Capelle, B. (2010). Doubler-upper nouns: A challenge for usage-based models of language? In A. Onysko & S. Michel (Eds.), *Cognitive Perspectives on Word Formation* (pp. 335-374). Berlin: Mounton de Gruyter.

Cappelle, B. (2014, November). [Review of the book Phrasal verbs: The English verb-particle construction and its history, by S. Thim]. *English Language and Linguistics, 18,* 572-586.

Cappelle, B., Shtyrov, Y. & Pulvermuller, F. (2010). Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain & Language, 115,* 189-201.

Carbary, K. M., Frohning, E. E., & Tanenhaus, M. K. (2010). Context, syntactic priming, and referential form in an interactive dialogue task: Implications for models of alignment. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 32*(32), 109-114..

Carlsen, C. (2012). Proficiency level: A fuzzy variable in computer learner corpora. *Applied Linguistics, 33*(2), 161-183.

Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course*. Boston, Massachusetts: Heinle & Heinle.

Chapman, D. (2008). Fixer-uppers and passers-by: Nominalization of verb-particle constructions. In S. M. Fitzmaurice & D. Minkova (Eds.), *Empirical and Analytical Advances in the Study of English Language Change* (pp. 265-300). Berlin: Mouton de Gruyter.

Chen, J. (2008). The acquisition of verb compounding in Mandarin Chinese (Unpublished doctoral dissertation). Vrije Universiteit, Amsterdam.

Chen, M. (2013). Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics, 18*(3), 418-442. doi: 10.1075/ijcl.18.3.07che

Chen, M. (2018). Beyond frequencies: Investigating the semantic and stylistic features of phrasal verbs in a three-year longitudinal study corpus by Chinese university students. In V. Brezina & L. Flowerdew (Eds.), *Learner corpus research: New perspectives and applications* (pp. 134-154). London: Bloomsbury.

Chen, X. & Smakman, D. (2014). Avoidance of phrasal verbs by learners of English: Definitional and methodological issues. Paper presented at the Een colloquium over universitair taalvaardigheidsonderwijs, Leiden University, The Netherlands. Retrieved from https://openaccess.leidenuniv.nl/bitstream/handle/1887/38548/05.pdf?sequence=1

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in cognitive science*, *9*(3), 542-551.

Claridge, C. (2000). *Multi-word verbs in early Modern English: A corpus-based study.* Amsterdam: Rodopi.

Columbus, G. (2013). In support of multiword unit classifications: Corpus and human rating data validate phraseological classifications of three different multiword types. *Yearbook of Phraseology*, *4*(1), 23-44.

Condon, N. (2008). How cognitive linguistic motivations influence the learning of phrasal verbs. In F. Boers & S. Lindstromberg (Eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 133-158). Berlin: Mouton de Gruyter.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied linguistics*, *29*(1), 72-89.

Conklin, K. & Carrol, G. (2019). First language influence on the processing of formulaic language in a second language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 62-77). Oxon: Routledge.

Consigny, A. (2001). *The semantics of phrasal verbs in English: A data-driven study* (Unpublished PhD thesis). University of Liverpool, UK.

Corder, S. (1971). Idiosyncratic dialect and error analysis. *International Review of Applied Linguistics in Language Teaching*, *9*(2), 147-160.

Corder, S. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.

Cowie, A. (1993). Getting to grips with phrasal verbs? *English Today 36, 9*(4), 38-41.

Cowie, A. P. (1998). Introduction. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 1-20). Oxford: Oxford University Press.

Cowie, A.P., & Mackin, R. (1975). *Oxford dictionary of current idiomatic English. Volume I: Verbs with prepositions & particles*. Oxford: Oxford University Press.

Cowie, A. P. & Mackin, R. (1993). *Oxford Dictionary of Phrasal Verbs*. Oxford: Oxford University Press.

Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning, 60*(3), 573-605.

Crossley, S., Salsbury, T., & Mcnamara, D. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics, 36*(5), 570-590.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System, 26*(2), 163–74.

Dagut, M. & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition, 7*(1), 73-79.

Danchev, A. (1992). The evidence of analytic and synthetic developments in English. In M. Rissanen, O. Ihalainen, T. Nevalainen, & I. Taavitsainen (Eds.), *History of Englishes* (pp. 25-41). Berlin: Mouton de Gruyter.

Darwin, C. M. & Gray, L. S. (1999). Going after the phrasal verb: An alternative approach to classification. *TESOL Quarterly, 33*(1), 65-83.

Davies, M. (2008). *The corpus of contemporary American English: 450 million words, 1990-present*. Provo, UT: Brigham Young University.

Davies, M. (2013). *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE*). Available online at https://corpus.byu.edu/glowbe/.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL), 2*, 225-246.

De Cock, S. (2005). Phrasal verbs and learners: How well do they get on? In M. Rundell (Ed.), *Macmillan phrasal verbs* plus (pp. LS16-20). Oxford: Macmillan.

Dehé, N. (2002). *Particle verbs in English: Syntax, information structure and intonation*. Amsterdam: John Benjamins.

Dehé, N. (2015). Particle verbs in Germanic. In P. Müller et al. (Eds.), *An international handbook of the languages of Europe. Vol. 1* (pp. 611-626). Berlin: Mouton de Gruyter.

Denison, D. (1981). *Aspects of the history of English group verbs with particular attention to the syntax of the Ormulum (Doctoral dissertation). Oxford University.*

Denison, D. (1984). The origins of the completive *up* in English. *Neuphilologische Mitteilungen, 86*, 37-61.

Denison, D. (1985). Why Old English had no prepositional passive. *English Studies, 66*, 189-204.

Denison, D. (1998). Syntax. In S. Romaine (Ed.), *The Cambridge History of the English Language* (Vol. 4, pp. 92-329). Cambridge: Cambridge University Press.

Deshors, S. C. (2016). Inside phrasal verb constructions: A co-varying collexeme analysis of verb-particle combinations in EFL and their semantic associations. *International Journal of Learner Corpus Research, 2*(1), 1-30.

Deshors, S. C., Götz, S., & Laporte, S. (2016). Linguistic innovations in EFL and ESL: Rethinking the linguistic creativity of non-native English speakers. *International Journal of Learner Corpus Research, 2*(2), 131-150.

Dewaele, J. M. (1998). Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics, 19*(4), 471-90.

Di Sciullo, A. M. & Williams, E. (1987). *On the definition of word*. Cambridge, MA: The MIT Press.

Díaz-Negrillo, A. & Domínguez, J. F. (2006). Error tagging systems for learner corpora. *Revista Española de Lingüística Aplicada, 19*, 83-102.

Diemer, S. (2014). What happened to the English prefix and could it stage a comeback? In K. Davidse, C. Gentens, L. Ghesquiere, & L. Vandelanotte (Eds.), *Corpus interrogation and grammatical patterns* (pp. 35-55). Amsterdam: John Benjamins.

Dirven, R. (2001). English phrasal verbs: Theory and didactic application. In M. Pütz, Niemeier, S., & Dirven, R. (Eds.), *Applied cognitive linguistics II: Language pedagogy* (pp. 3-27). Berlin: Mouton de Gruyter.

Dixon, R. M. W. (1982). The grammar of English phrasal verbs. *Australian Journal of Linguistics, 2*, 1-42.

Dóczi, B. & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. New York: Oxford University Press.

Dufresne, M., Dupuis, F. & Tremblay, M. (2003). Preverbs and particles in Old French. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 2003* (pp. 33-60). Dordrecht: Kluwer.

Edwards, A., & Laporte, S. (2015). Outer and expanding circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide, 36*(2), 135-169.

El-Dakhs, D. A. S. (2016). The lexical knowledge and avoidance of phrasal verbs: The case of Egyptian learners of English. *International Journal of Applied Linguistics and English Literature, 5*(1), 132-144.

Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 219-238). Amsterdam: John Benjamins.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194.

Ellis, N. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics, 32*, 17-44.

Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375-396.

Erman, B., Forsberg Lundell, F., Lewis, M. (2016). Formulaic language in advanced second language acquisition and use. In K. Hyltenstam (Ed.), *Advanced proficiency and exceptional ability in second languages* (pp. 111-147). Berlin: Walter de Gruyter.

Eskildsen, S. W. (2016). 11 The emergence of creativity in L2 English: A usage-based case-study. In N. Bell (Ed.), *Multiple perspectives on language play* (pp. 281-316). Berlin: Mouton de Gruyter.

Farrell, P. (2005). English verb-preposition constructions: Constituency and order. *Language*, *80*(1), 96-137.

Fernando, C. (1996). *Idioms and idiomaticity*. Oxford: Oxford University Press.

Fletcher, B. (2005). Registers and phrasal verbs. In M. Rundell (Ed.), Macmillan phrasal verbs plus (pp. LS13-15). Oxford: Macmillan.

Folse, K. S. (2004). Myths about teaching and learning second language vocabulary: What recent research says. *TESL Reporter, 37*(2), 1-13.

Foster, P. & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299-323.

Fraser, B. (1976). *The verb-particle combination in English*. New York: Academic Press.

Gablasova, D. (2017). *Trinity Lancaster Corpus sample*. Unpublished manuscript, Department of Linguistics and English Language, Lancaster University, Lancaster, United Kingdom.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning*, *67*(S1), 130-154.

Gablasova, D., Brezina, V. & McEnery, T. (forthcoming). The Trinity Lancaster Corpus: Development, description and application. *International Journal of Learner Corpus Research,* 5(2).

Gardner, D. & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly, 41*(2), 339-359.

Gardner, D., & Davies, M. (2018). Sorting them all out: Exploring the separable phrasal verbs of English. *System, 76*, 197-209.

Garnier, M. & Schmitt, N. (2015). The PHaVE list: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research, 19*(6), 645-666.

Garnier, M. & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System, 59*, 29-44.

Garside, R. (1987). The CLAWS word-tagging system. In R. Garside, G. Leech, & G. Sampson (Eds.), *The computational analysis of English: A corpus-based approach*. London: Longman.

Gass, S. M., & Mackey, A. (2007). Input, interaction and output: An overview. *AILA review, 19*(1), 3-17.

Gilquin, G. (2011). Corpus linguistics to bridge the gap between world Englishes and learner Englishes. *Communicación en el siglo XXI, 2*, 638-642.

Gilquin, G. (2015a). At the interface of contact linguistics and second language acquisition research: New Englishes and learner Englishes compared. *English World-Wide, 36*(1), 91-124.

Gilquin, G. (201b5). The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory, 11*(1), 51-88.

Gilquin, G., De Cock, S., & Granger, S. (2010). *Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-laNeuve: Presses universitaires de Louvain. .

Giora, R. (2003). *On our mind: salience, context and figurative language*. Oxford: Oxford University Press .

Glodovic, A. (2014). Function of the adverbial particle out in English phrasal verbs: Corpus analysis. *V2014*, 9, 52-62.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, *7*(5), 219-224.

González Fernández, M. J. (1997). Sobre la motivación semántica de las expresiones pleonásticas de movimiento: subir arriba, bajar abajo, entrar adentro y salir afuera. In C. Company (Ed.), *Cambios diacrónicos en el español* (pp. 123-141). M*éxico: Universidad Nacional Autónoma de México.*

González, R. A. (2010). L2 Spanish acquisition of English phrasal verbs: A cognitive linguistic analysis of L1 influence. In M. Campoy, B. Belles-Fortuno, & M. Gea-Valor, M. (Eds), *Corpus-based approaches to English language teaching* (pp. 149-166). London: Continuum.

Götz, S. (2013). *Fluency in native and non-native English speech* (Vol. 53). Amsterdam: John Benjamins.

Gouverneur, C. (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 & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 223-246). Amsterdam: John Benjamins.

Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in Contrast* (pp. 37-51). Lund: Lund University Press.

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

Granger, S. (1998b). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.

Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal, 20*(3), 465-480.

Granger, S. (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* (pp. 13-32). Amsterdam: John Benjamins.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research 1(*1), 7-24 .

Granger, S. (2019). Formulaic sequences in learner corpora: Collocations and lexical bundles. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 228-247). Oxon: Routledge.

Granger, S., Paquot, M. & Rayson, P. (2006). Extraction of multi-word units from EFL and native English corpora. The phraseology of the verb 'make'. In A. H. Buhofer & H. Burger (Eds.), *Phraseology in Motion I: Methoden und Kritik (pp. 57-68).* Schneider Verlag.

Granger, S. & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27-49). Amsterdam: John Benjamins Publishing Company.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International Corpus of Learner English. Handbook and CD-ROM*. Version 2. Louvain-laNeuve: Presses universitaires de Louvain.

Grant, L. & Bauer, L. (2004). Criteria for redefining idioms: Are we barking up the wrong tree? *Applied Linguistics, 25*(1), 38-61.

Greenbaum, S. (1991). ICE: The International Corpus of English. *English Today, 7*, 3-7.

Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement*. New York: Continuum.

Gries, S. (2015). Polysemy. In E. Dąbrowska & S. D. Dagmar (Eds.), *Handbook of Cognitive Linguistics* (pp. 472-490). Berlin & Boston: De Gruyter Mouton.

Hampe, B. (2002). Superlative verbs: *A corpus-based study of semantic redundancy in English verb-particle constructions (Vol. 24)*. Tübingen: Gunter Narr Verlag.

Hasselgård, H. & Johansson, S. (2011). Learner corpora and contrastive interlanguage analysis. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora. In honour of Sylviane Granger* (pp. 33-61). Amsterdam: John Benjamins.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*, 237–260.

Hijazo-Gascón, A., & Ibarretxe-Antuñano, I. (2013). Las lenguas románicas y la tipología de los eventos de movimiento. *Romanische Forschungen, 125*(4), 467-494.

Hiltunen, R. (1983). *The decline of the prefixes and beginnings of the English phrasal verb: The evidence from some Old and early Middle English texts*. Turku: Turun Yliopisto.

Hiltunen, R. (1994). On phrasal verbs in Early Modern English: Notes on lexis and style. In D. Kastovsky (Ed.), *Studies in Early Modern English* (pp. 129-140). Berlin: Mouton de Gruyter.

Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics, 41*(4), 667-683.

Hoekstra, J. F. (2001). Standard West Frisian. In H. Haider Munske (Ed.), *Handbook of Frisian studies* (pp. 83-98). Tübingen: Niemeyer.

Hoffmann, T., & Trousdale, G. (Eds.). (2013). *The Oxford handbook of construction grammar*. Oxford: Oxford University Press.

Huat, C. M. (2012) . Learner corpora and second language acquisition. In K. Hyland, C. M. Huat, & M. Handford (Eds.), *Corpus Applications in Applied linguistics* (pp. 191-207). London: Continuum.

Huddleston, R. & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hulstijn, J. H. & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition, 11*(3), 241–255.

Iacobini, C. (2009a). Phrasal verbs between syntax and lexicon. *Rivista di Linguistica, 21*(1), 97-117.

Iacobini, C. (2009b). The role of dialects in the emergence of Italian phrasal verbs. *Morphology, 19*, 15-44.

Iacobini, C. (2015). Particle verbs in Romance. In P. Müller et al. (Eds.), An international handbook of the languages of Europe. Vol. 1 (pp. 627-659). Berlin: Mouton de Gruyter.

Iacobini, C., & Masini, F. (2006). The emergence of verb-particle constructions in Italian: Locative and actional meanings. *Morphology, 16*(2), 155-188.

Irujo, S. (1993). Steering Clear: Avoidance in the Production of Idioms. *International Review of Applied Linguistics, 31*(3), 205-219.

Ishizaki, Y. (2012). A usage-based analysis of phrasal verbs in Early and Late Modern English. *English Language & Linguistics, 16*(2), 241-260.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. In N. Dehé, R. Jackendoff, A. McIntyre, A., & S. Urban (Eds.) *Verb-particle Explorations* (pp. 67-94). Berlin: Mouton de Gruyter.

James, C. (1998). *Errors in language learning and use : Exploring error analysis*. London: Routledge.

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in them interlanguage lexicon. *Language Learning, 50*(2), 245-309.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York: Routledge.

Kachru, B. B. (1982). Models for non-native Englishes. In B. B. Kachru (Ed.), *The Other Tongue: English across cultures* (pp. 48-74). Urbana and Chicago: University of Illinois Press.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.

Kamarudin, R. (2013). *A study on the use of phrasal verbs by Malaysian learners of English* (Unpublished doctoral dissertation). University of Birmingham, UK.

Kapatsinski, V. & Radicke, J. (2009). Frequency and the emergence of prefabs: Evidence from monitoring. In C. Roberta, E. A. Moravcsik, H. Ouali, & K. M. Wheatley (Eds.), *Formulaic language* (pp. 499-520). Amsterdam: John Benjamins Publishing Company.

Karimi, S. N. (2017). Effects of output task types on the acquisition of English phrasal verbs. *MJAL, 9*(2), 123-150.

Kaszubski, P. (2000). *Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: A contrastive, corpus-based approach*. http://main.amu.edu.pl/~przemka/rsearch.html

Ke, Y. (2013). The phraseology of phrasal verbs in English: A corpus study of the language of Chinese learners and native English writers (Unpublished doctoral dissertation). University of Birmingham, UK.

Ke, Y. (2017). A bi-axis model for profiling English phrasal verbs for pedagogic purposes. *TESOL Quarterly, 51*(4), 972-984.

Kecskes, I. (2006). On my mind: Thoughts about salience, context and figurative language from a second language perspective. *Second Language Research, 22*(2), 219-237.

Khabbazbashi, Nahal. (2017). Topic and background knowledge effects on performance in speaking assessment. *Language Testing, 34*(1), 23-48.

Kharitonova, A. (2013). *Lexical transfer and avoidance in the acquisition of English phrasal verbs: Evidence from Russian and Norwegian learners of English* (Unpublished master's thesis). University of Oslo, Sweden.

Kilgarriff, A.; Rychly, P.; Smrz, P. & Tugwell, D. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1–30.

Kim, S. N., & Baldwin, T. (2010). How to pick out token instances of English verb-particle constructions. *Language resources and evaluation, 44*(1-2), 97-113.

Kroch, A., Santorini, B., & Diertani, A. (2010). Penn parsed corpus of Modern British English (PPCMBE). Retrieved from http://www.ling.upenn.edu/hist-corpora/PPCMBE-RELEASE-1/index.html.

Kurtyka, A. (2001). Teaching English phrasal verbs: A cognitive approach. In M. Pütz, Niemeier, S., & Dirven, R. (Eds.), *Applied cognitive linguistics II: Language pedagogy* (pp. 29-54). Berlin: Mouton de Gruyter.

Kytö, M. & Smitterberg, E. (2006). 19th-century English: An age of stability or a period of change? In R. Facchinetti & M. Rissanen (Eds.), *Corpus-based studies of diachronic English* (pp. 199-230). Bern: Peter Lang.

Ladányi, M. (2015). Particle verbs in Hungarian. In P. Müller et al. (Eds.), An international handbook of the languages of Europe. Vol. 1 (pp. 660-672). Berlin: Mouton de Gruyter.

Larsen-Freeman, D. (2002). Making sense of frequency. *Studies in Second Language Acquisition, 24*(2), 275-285.

Larsen-Freeman, D. (2014). Another step to be taken: Rethinking the end point of the interlanguage continuum. In Z. Han & E. Tarone (Eds.), *Interlanguage: Forty years later* (pp. 203-220). Amsterdam: John Benjamins.

Laufer, B. & Eliasson, S. (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition, 15*(1), 35-48.

Lee, J. Y. (2015). The use of English phrasal verbs in American spoken corpora. *International Journal of Language Studies, 9*(2), 27-48.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Longman.

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

Leone, L. (2016). Phrasal verbs and analogical generalization in Late Modern spoken English. *ICAME Journal, 40*(1), 39-62.

Li, W., Zhang, X., Niu, C., Jiang, Y., & Srihari, R. (2003). An expert lexicon approach to identifying English phrasal verbs. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1* (pp. 513-520). Association for Computational Linguistics.

Liao, Y. & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, 54(2), 193–226.

Lin, J. (2015). The encoding of motion events in Mandarin Chinese. In W. S. Y. Wang & C. Sun (Eds.), *The Oxford handbook of Chinese linguistics* (pp. 322-335). New York: Oxford University Press.

Lindner, S. J. (1983). *A lexico-semantic analysis of English verb particle constructions*. Bloomington, Indiana: Indiana University Linguistics Club.

Lindstromberg, S. (2010). *English prepositions explained*. Amsterdam: John Benjamins.

Lipka, L. (1972). *Semantic structure and word-formation: Verb-particle constructions in contemporary English*. München: Wilhelm Fink Verlag.

Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *TESOL Quarterly, 45*, 661-668.

Liu, D., & Myers, D. (2018). The most-common phrasal verbs with their key meanings for spoken and academic written English: A corpus analysis. *Language Teaching Research*, 1-18.

Loewen, S., & Plonsky, L. (2016). *An A–Z of applied linguistics research methods*. New York: Palgrave.

Love, R., Dembry, C., Hardie, A., Brezina, V. and McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics, 22*(3): 319-344.

Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning, 65*(1), 63-88.

Lüdeling, A. & Hirschmann, H. (2015). Error annotation. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 135-158). Cambridge: Cambridge University Press.

Lüdeling, A., Hirschmann, H., & Shadrova, A. (2017). Linguistic models, acquisition theories, and learner corpora: Morphological productivity in SLA research exemplified by complex verbs in German. *Language Learning, 67*(S1), 96-129.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied linguistics, 27*(3), 405-430.

Mahmoud, A. (2015). Under-representation of phrasal verbs in EFL university students' free writing. *International Journal of English and Education, 4*(1), 261-268.

Mahpeykar, N., & Tyler, A. (2015). A principled cognitive linguistics account of English phrasal verbs with up and out. *Language and Cognition, 7*(1), 1-35.

Maiden, M., & Robustelli, C. (2014). *A reference grammar of modern Italian*. London: Routledge.

Makai, A. (1972). *Idiom structure in English*. The Hague: Mouton.

Märzinger, K. (2013). *The use of phrasal verbs in English as a Lingua Franca: An analysis of authentic EFL data from VOICE*. Akademiker Verlag.

Massini, F. (2005). Multi-word expressions between syntax and the lexicon: The case of Italian verb-particle constructions. *SKY Journal of Linguistics, 18,* 145-173.

Masini, F. (2008). Verbi sintagmatici e ordine delle parole. In M. Cini (ed.), *I verbi sintagmatici in italiano e nelle varieta dialettali. Stato dell'arte e prospettive di ricerca* (pp. 83-102). Frankfurt: Peter Lang.

Mateu, J., & Rigau, G. (2010). Verb-particle constructions in Romance: A lexical-syntactic account. *Probus, 22*(2), 241-269.

Matlock, T. & Heredia, R. R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 251-274). The Netherlands: Elsevier.

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

McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice.* Cambridge: Cambridge University Press.

Meunier, F. (2015). Developmental patterns in learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp.379-400). Cambridge: Cambridge University Press.

Meunier, F. & Granger, S. (2008). *Phraseology in Foreign Language Learning and Teaching.* Amsterdam: John Benjamins.

Mitchell, T. F. (1958). Syntagmatic relations in linguistic analysis. *Transactions of the Philological Society*, 101-118.

Mondor, M. (2008). *Figuring it out. A corpus-based comparison of the verb-particle construction in argumentative writing by Swedish advanced learners and native speakers of English* (Unpublished doctoral dissertation). University of Gothenburg, Sweden.

Moon, R. (1998). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 79-100). Oxford: Clarendon Press.

Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2: A basis for creative construction? *Studies in Second Language Acquisition, 21*(1), 49-80.

Nassaji, H., & Tian, J. (2010). Collaborative and individual output tasks and their effects on learning English phrasal verbs. *Language Teaching Research, 14*(4), 397-419.

Nation, I. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA.: Heinle.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Newton, J. & Graeme, K. (1996). Effects of communication tasks on the grammatical relations marked by second language learners. *System, 24*(3), 309-322.

Nurmi, A. (1999). The corpus of Early English correspondence sampler (CEECS). *ICAME Journal, 23*, 53-64.

O'Dowd, E. M. (1998). *Prepositions and particles in English: A discourse-functional account*. New York: Oxford University Press.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Odlin, T. (2003). Cross-linguistic influence. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 436-486). Malden, MA: Blackwell Publishing.

Ortega, L. Byrnes, H. (Eds). *The longitudinal study of advanced L2 capacities*. London: Routledge.

Palmer, F. R. (1988). *The English verb*. London: Longman.

Paquot, M. & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics, 32,* 130-149.

Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics, 18*(3), 391-417.

Pawley, A. & Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C. Richards & R.W. Schmidt (Eds.), *Language and communication* (pp.191-226). New York: Longman.

Quirk, R.; Greenbaum, S.; Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Reznicek, M., Lüdeling, A., & Hirschmann, H. (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* (pp. 101-123). Amsterdam: John Benjamins.

Ringbom, H. (2016). Comprehension, learning and production of foreign languages: The role of transfer. In Alonso Alonso, R. (Ed.), *Crosslinguistic influence in second language acquisition* (pp. 38-52). Bristol: Multilingual Matters.

Rodríguez-Puente, P. (2012). The development of non-compositional meanings in phrasal verbs: A corpus-based study. *English Studies, 93*(1), 71-90.

Rodríguez-Puente, P. (2016). Tracking down phrasal verbs in the spoken language of the past: Late Modern English in focus. *English Language and Linguistics*, 1-29.

Rodríguez Puente, P. (2019). *The English phrasal verb, 1650-present: History, stylistic drifts, and lexicalisation*. Cambridge: Cambridge University Press.

Rudzka-Ostyn, B. (1988). Semantic extensions into the domain of verbal communication. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics* (pp. 507-554). Amsterdam: John Benjamins.

Saiya, C. (2011). *Avoidance of phrasal verbs by M.6 students at Thiamudomsuksapattanakarn Nonthaburi School* (Unpublished master's thesis). Thammasat University, Thailand.

Samuda, V. & Bygate, M. (2008). *Tasks in second language learning*. New York: Palgrave Macmillan.

Sara, H. & Mohammadreza, T. (2013). Study on avoidance behaviour among Persian EFL learners: Phrasal verbs in focus. *Greener Journal of Educational Research, 3*(6), 238-248.

Schmitt, N. & Carter, R. (2004). Formulaic sequences in action. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1-22). Benjamins: Amsterdam.

Schmitt, N. & Redwood, S. (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A Taste for Corpora* (pp. 173-207). Amsterdam: John Benjamins.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12(*3), 329-363.

Schneider, E. W. (2014). New reflections on the evolutionary dynamics of world Englishes. *World Englishes, 33*(1), 9-32.

Schneider, G. & Zipp, L. (2013). Discovering new verb-preposition combinations in New Englishes. *Studies in Variation, Contacts and Change in English, 13*. Retrieved from http://www.helsinki.fi/varieng/series/volumes/13/schneider_zipp.pdf

Schneider, G., & Gilquin, G. (2016). Detecting innovations in a parsed corpus of learner English. *International Journal of Learner Corpus Research, 2(*2), 177-204.

Selinker, L. (1992). *Rediscovering Interlanguage*. London: Longman.

Side, R. (1990), Phrasal verbs: Sorting them out. *ELT Journal, 44*(2), 144-152.

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan corpus of academic spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (2004). Trust the text. In J. Sinclair (Ed.), *Trust the Text: Language, Corpus and Discourse* (pp. 9-23) London: Routledge.

Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory, 11*(2), 285-301.

Siyanova, A. & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics in Language Teaching*, *45*, 119-139.

Siyanova-Chanturia, A. & Nation, I.S.P. (2017). Teaching communicative vocabulary. In H.P. Widodo, A. Wood, D. Gupta, & W. Cheng (Eds.), *Asian English Language Classrooms: Where Theory and Practice Meet* (pp. 98-112). New York: Routledge.

Siyanova-Chantura, A. & Pellicer-Sánchez, A. (2019). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 1-15). Oxon: Routledge.

Siyanova-Chantura, A. & Van Lancker Sidtis, D. (2019). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 38-61). Oxon: Routledge.

Sjoholm, K. (1995). The Influence of crosslinguistic, semantic, and input factors on the acquisition of English phrasal verbs: A comparison between Finnish and Swedish learners at an intermediate and advanced level. Åbo, Finland: Åbo Akademi University Press.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Slobin, D. I. (2004). The many ways to search for a frog. In S. Stromqvist & L. Verhoeven (Eds.), Relating events in narrative: Typological and contextual perspectives (pp. 219-257). Mahwah, NJ: Lawrence Erlbaum.

Smitterberg, E. (2008). The progressive and phrasal verbs: Evidence of colloquialization in nineteenth-century English? In T. Nevalainen, I. Taavitsainen, P. Pahta, & Korhonen, M. (Eds.), *The dynamics of linguistic*

*variation: Corpus evidence on English past and present* (pp. 269-290). Amsterdam: John Benjamins.

Sroka, K. A. (1972). *The syntax of English phrasal verbs*. The Hague: Mouton.

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in second language acquisition, 28*(3), 387-422.

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. London: Routledge.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description (Vol. 3: Grammatical categories and the lexicon)* (pp. 36-149). Cambridge, UK: Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics. Vol.II: Typology and process in concept structuring*. Cambridge, MA: MIT Press.

Thim, S. (2012). *Phrasal verbs: The English verb-particle construction and its history*. Germany: Walter de Gruyter.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins Publishing.

Torres Martínez, S. (2015). A constructionist approach to the teaching of phrasal verbs: Dispelling the verb+ particle myth in multiword verb instruction. *English Today, 31*(3), 46-58.

Tracy-Ventura, N. & Myles, F. (2015). The importance of task variability in the design of learner corpora for SLA research. *International Journal of Learner Corpus Research, 1*(1), 58-95.

Trinity College London. (2018). *Exam information: Graded examinations in spoken English (GESE)* (online ed.). Trinity College London.

Trinity College London. (c2019). *Graded Examinations in Spoken English (GESE).* Retrieved from http://www.trinitycollege.com/site/?id=368

Van Patten, B. & Williams, J. (2015). *Early theories in SLA.* In B. Van Patten & J. Williams. (Eds.), *Theories in second language acquisition: An introduction* (pp. 17-33). London: Routledge.

Van Rooy, B. (2011). A principled distinction between error and conventionalized innovation in African Englishes. In Mukherjee, J. & Hundt, M. (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 189-207). Amsterdam: John Benjamins.

Verspoor, M. H., & Lowie, W. M. (2003). Making sense of polysemous words. *Language Learning, 53*(3), 547-586.

Verspoor, M., Lowie, W., Chan, H., & Vahtrick, L. (2017). Linguistic complexity in second language development: Variability and variation at advanced stages. *Recherches en didactique des langues et des culture . Les cahiers de l'Acedle, 14*(1).

Viberg, Å. (1996). Crosslinguistic lexicology: The case of English go and Swedish ga. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in Contrast* (pp. 151-182). Lund: Lund University Press.

Viberg, Å. (2002). Polysemy and disambiguation cues across languages: The case of Swedish fa and English get. In B. Altenberg & S. Granger (Eds.), *Lexis in contrast: Corpus-based approaches* (pp. 119-150). Amsterdam: John Benjamins.

Villavicencio, A. (2003). Verb-particle constructions and lexical resources. In *Proceedings of the ACL 2003 workshop on multiword expressions: Analysis, acquisition and treatment-Volume 18* (pp. 57-64). Association for Computational Linguistics.

VOICE. (2013). *The Vienna-Oxford International Corpus of English (version 2.0 XML)*. Director: Barbara Seidlhofer; Researchers: Angelika Breiteneder, Theresa Klimpfinger, Stefan Majewski, Ruth Osimk-Teasdale, Marie-Luise Pitzl, Michael Radeka.

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *Modern Language Journal, 96*(4), 576-598.

Waibel, B. (2008). *Phrasal verbs: German and Italian learners of English compared.* Saarbrücken: VDM Verlag Dr. Müller.

Webb, S. (2018). Advanced lexical development. In P. A. Malovrh & A. G. Benati (Eds.), *The handbook of advanced proficiency in second language acquisition* (pp. 401-418). United States: Wiley.

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

White, B. J. (2012). A conceptual approach to the instruction of phrasal verbs. *The Modern Language Journal, 96*(3), 419-438.

Wierszycka, J. (2013). Phrasal verbs in learner English: A semantic approach. *Research in Corpus Linguistics, 1*, 81-93.

Wild, K. (2010). *Attitudes towards English usage in the Late Modern period: The case of phrasal verbs* (Unpublished doctoral dissertation). University of Glasgow, UK.

Wild, K. (2011). Phrasal verbs: 'A process of the common, relatively uneducated, mind'? *English Today, 27*(4), 53-57.

Williams, J. R., Lessard, P. R., Desu, S., Clark, E. M., Bagrow, J. P., Danforth, C. M., & Dodds, P. S. (2015). Zipf's law holds for phrases, not words. *Scientific reports, 5*, 12209.

Wood, D. (2002). Formulaic language acquisition and production: Implications for teaching. *TESL Canada Journal, 20*(1), 1-15.

Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction*. London: Bloomsbury.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford University Press.

Wray, A. (2019). Concluding question: Why don't second language learners more proactively target formulaic sequences? In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 248-269). Oxon: Routledge.

Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. London: Continuum.

Wulff, S. (2019). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 19-37). Oxon: Routledge.

Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *Tesol Quarterly, 44*(2), 250-273.

Yoon, H. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System, 66,* 130-141.

Yorio, C. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss* (pp. 55-72). Cambridge: Cambridge University Press.

You, Y. S. (1999). Avoidance phenomena of phrasal verbs by Korean learners of English. *English Teaching*, *54*(3), 135-155.

Zhang, X., & Wen, J. (2019). Exploring multiple constraints on second language development of English polysemous phrasal verbs. *Applied Psycholinguistics*, 1-29.

Zhang, Y & Rao, Q. (2012). Jiyu hanying pingxing yuliaoku de yingyu duanyu dongci hanyi yanjiu. [A parallel corpus-based study on Chinese translation of English phrasal verbs. *Journal of Hubei University (Philosophy and Social Science), 39*(5).

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.

Zipp, L., & Bernaisch, T. (2012). Particle verbs across first and second language varieties of English. In M. Hundt & U. Gut (Eds.), *Mapping unity in diversity world-wide: Corpus-based studies of new Englishes* (pp. 167-196). Amsterdam: John Benjamins.