

The Passive and the Lexis-Grammar Interface:

An Inter-varietal Perspective

Gaëtanelle Gilquin and Sylviane Granger

1 Introduction

The passive has traditionally been seen as a purely grammatical phenomenon, resulting from the transformation of an active sentence. From a foreign/second language (L2) perspective, it is a complex grammatical structure involving morphology and syntax as well as aspects related to text organization and register. For L2 learners, the difficulty is compounded by the fact that the forms, meanings and contexts of use of the passive differ greatly across languages. As a result, the passive has been shown to pose great difficulties to L2 learners, both in comprehension and production.

Unlike traditional approaches, corpus-based studies of the passive have shown that the passive is not a purely grammatical phenomenon. First, the distinction between the active and the passive is not a strict dichotomy but a gradient ranging from fully verbal to fully adjectival. Second, the passive displays lexical effects, with some verbs being more attracted to the passive voice than others and/or typically used in a limited set of phraseological units. Yet, this view is still largely absent from the field of second language acquisition. In this chapter, we adopt the view of the passive as a lexico-grammatical phenomenon to study different varieties of L2 English as produced by learners from eight mother tongue (L1) backgrounds, representing both

English as a foreign language (EFL) and English as a second language (ESL). The focus is on BE-passives only.

The chapter is structured as follows. Section 2 describes the passive as a multifaceted construction and shows what aspects of the passive are dealt with in reference and instructional materials. Section 3 provides an overview of L2 studies of the passive, with a focus on errors, under-use, and lexical and phraseological patterning. In Section 4 we consider the continuum between EFL and ESL underlying our choice of L2 varieties, while in Section 5 we describe our data and methodology. The results are presented in Section 6, which examines the frequency of the passive, and in Section 7, which looks at the lexis-grammar interface, and, more precisely, the passive ratio and phraseological sequences. Section 8 concludes the chapter.

2 The Passive: A Multifaceted Construction

Whilst essentially considered as a grammatical structure, the passive is the result of the interplay between a range of other factors. In this section we first present the aspects that have traditionally been described – albeit to varying degrees – in reference and ELT (English language teaching) grammars, i.e. grammar, discourse and register. We then turn to the more recently uncovered lexical and phraseological aspects, which have yet to make their way into reference and instructional materials. In the third subsection we address the passive gradient, i.e. the categorization of passive constructions along a continuum ranging from the most central representatives to the most peripheral.

2.1 Grammar, discourse and register

The passive has pride of place in all reference grammars of English. It is usually presented as a structural counterpart of the more basic active construction. For Huddleston and

Pullum (2002: 46), the passive construction can be described derivatively, i.e. in terms of how it differs from the basic, canonical construction. As illustrated in (1), the active-passive transformation involves the addition of the BE auxiliary (or less frequently GET) and the past participle morpheme of the main verb, and re-ordering of the object and subject. The object becomes the subject of the passive sentence. The subject is shifted to post-verbal position and becomes a prepositional phrase introduced by the preposition *by* (or occasionally some other preposition); in many cases, it is simply omitted (agentless passive).

(1) a. Pat solved the problem.

b. The problem was solved by Pat. (Huddleston & Pullum, 2002: 46)

The active and passive constructions are syntactically different but their propositional meanings are essentially the same. The main problem is therefore to know when to use either structure, i.e. to be aware of their respective discourse functions. One of the essential factors to consider in this connection is that the passive construction is an ‘information-packaging construction’ (Huddleston & Pullum, 2005: 242), which can make it possible to re-establish the normal given-new order in the sentence, as illustrated by (2), where the active (2a) has a new-information subject and the passive (2b) a more typical old-information subject.

(2) a. A dog attacked me in the park.

b. I was attacked by a dog in the park. (Huddleston & Pullum, 2005: 242)

Other factors involve thematic coherence, the principle of end weight and, for the agentless passive, omission of the agent because it is unknown, redundant or general (Biber *et al.*, 1999: 938-942; Huddleston & Pullum, 2002: 1443-1447).

Another factor which tips the balance to the use of the active or passive voice is the register of the text: some registers favour the passive, others disfavour it. This aspect is alluded to in most grammars but is especially in evidence in corpus-based grammars, most notably Biber *et al.* (1999), which compares the frequency of the passive in four genres – conversation, fiction, news and academic prose. The analysis shows that the passive is most frequent in academic prose, where it accounts for 25% of all finite verbs, and least frequent in conversation, where it accounts for only 2%; it is also common in news (15%). Several studies of academic texts written by both L1 and L2 writers have demonstrated that the passive is a linguistic feature that is positively associated with success in quantitative writing assessment (Wilson, 2006).

Arguably ‘one of the thorniest problems in L2 grammar instruction’ (Hinkel, 2002: 233), the passive is also a key focus in ELT grammars and textbooks. However, as shown by a survey of ELT grammars (Granger, 2013), the sections on the passive are biased towards the structural aspects, to the detriment of aspects pertaining to functions and contexts of use. Note, however, that Larsen-Freeman and Celce-Murcia’s (2016: 352) grammar is a notable exception as the authors explicitly state that, while most EFL/ESL grammars seem to assume that the greatest challenge for learners is the form of the passive, their perspective is different: ‘Our experience has shown us that it is learning when to use the English passive that presents the greatest long-term challenge to ESL/EFL students’. On the whole, however, the bulk of the sections in ELT grammars is devoted to passive morphology (addition of BE and past participle morpheme), the different passive clause types (monotransitive, ditransitive, impersonal, etc.) and issues related to aspect, tense and modality. The main focus is on the active-passive transformation of decontextualized sentences, regularly resulting in highly unidiomatic sentences (such as *That John couldn’t possibly win was recognized by everyone* in Cowan, 2008: 394). A great deal of attention is also directed to constraints on the passive, in particular the issue of unpassivizable

verbs. Not all transitive verbs can be passivized: verbs like *have*, *cost*, *equal*, *last* or *resemble* do not normally have corresponding passive forms (**Two houses are had by John*). This observation is often mistakenly extended to the whole category of stative verbs (Cowan, 2008: 400; Larsen-Freeman & Celce-Murcia, 2016: 356). This marked focus on passive constraints may lead learners to avoid the passive and is certainly not conducive to encouraging them to use it, which should be a major objective (Parrott, 2000: 287, 295). It is all the more detrimental as it reinforces the generally negative view of the passive found in many style guides and writing instruction manuals. For Garner (2016: 676), for example, one ought to have ‘a presumption against the passive’, the active having ‘palpable advantages in most contexts’ (see Pullum, 2014, for a rebuttal of stylistic charges against the passive).

2.2 Lexis and phraseology

The passive is not only a grammatical structure that makes it possible to reorganize discourse elements and that varies according to register: it is also a construction that has been shown to display strong lexical and phraseological preferences, although such aspects tend to be neglected in reference and instructional materials.

The main lexical aspects that are included in grammars are lists of verbs that cannot passivize. Verbs are presented as either passivizable or unpassivizable. However, corpus studies have shown that verbs display different degrees of passivizability. Biber *et al.* (1999: 478) provide lists of lexical verbs that have a very high passive frequency, including verbs that mostly occur in the passive in academic prose (e.g. BE *based (on)*, BE *situated*). Granger (2013) computed the passive ratio (i.e. the number of passive forms of a verb over the total number of occurrences of that verb) of the verbs used in the BNC Baby, a subset of the British National Corpus.¹ The

¹ <http://www.natcorp.ox.ac.uk/corpus/babyinfo.html>

results highlight considerable differences, with the verb OBLIGE, for example, displaying a ratio of 68.2% and the verb WANT a ratio as low as 0.8%. Using the method of collocation analysis, Gries and Stefanowitsch (2004) also underline the strong association of certain verbs with the passive and show that verbs attracted to the passive tend to belong to distinctive semantic classes. These results lead the authors to claim that the active-passive pair is not ‘a purely syntactic alternation’ (Gries & Stefanowitsch, 2004: 108) and that the ‘passive voice is a construction in its own right with its own specific semantics’ (2004: 110). ELT grammars have not yet taken into account this interdependence between grammar and lexis, which is one of the hallmarks of corpus linguistics. Some grammars provide general comments such as ‘Use of the passive is partly a matter of choice, though some verbs may be used more often in passive than active’ (Vince, 2008: 34), which are too vague to be helpful. It should be noted, however, that corpus-based resources have begun to integrate lexical aspects into their section on passives.

Going beyond the verbs that are preferred in the active vs passive voice, Lehmann and Schneider (2009) investigate the interaction between the voice of verbs and their subjects/objects. They highlight, among other things, typical subject-passive verb combinations in their corpus data (e.g. *shot* as a subject of *fired* or *breakfast* as a subject of *served*), as well as subject-verb combinations that are exclusively found in the passive (e.g. *hotel* and *situated* or *little* and *known*). Studies of lexical bundles have also frequently brought to light phraseological sequences that include a passive verb, especially in academic writing. Biber *et al.* (1999: 1019-1020), for example, describe patterns such as ‘anticipatory *it* + verb phrase’ where the verb is usually in the passive, as illustrated by *it can be seen* or *it should be noted*, and ‘passive verb + prepositional phrase fragment’, as in *are shown in table* or *is referred to as* (see also Hyland, 2008; Rowley-Jolivet, 2001). Such insights have not really made their way

into ELT grammars yet, although rare references can be found in some corpus-based ELT grammars: Conrad and Biber (2009: 46, 50) give examples of clusters made up of a verb in the passive and a preposition (*be associated with, can be interpreted as, be accompanied by*, etc.); Swales and Feak (2004: 120-121) mention the use of passive verbs in references to a visual (*illustrated in, seen from*, etc.); Bunting and Diniz (2013: 237) list verbs that frequently occur in the structure ‘*it + passive form of the present perfect*’ (e.g. *suggested, reported*) and modals that frequently occur with passive *it* constructions (*can, should, could* and *must*), as in *it can be argued that*.

2.3 The passive gradient

While it is customary to refer to ‘the passive’ as if it were a unitary structure, many grammarians and linguists point to the ambiguity of the BE + past participle (BE *Ved*) construction. As early as 1966, in the first fully corpus-based study of the passive in a range of mostly written registers, Svartvik identified no fewer than eight categories distributed along a continuum ranging from the most central passives which ‘have close systemic transformational relation with the active voice’ (Svartvik, 1966: 156) to less central ones which have ‘mixed verbal and adjectival properties’ (1966: 147) and/or for which ‘agent extension is unlikely or impossible’ (1966: 148). Examples (3) to (10) are examples given by Svartvik to illustrate the eight categories, ordered from the most central to the most peripheral, and specify the terms used to refer to each of them.

(3) He *was given* this puppy by a farmer in the Welsh hills [animate agent passive]

(4) The removal of large quantities of water (...) *is facilitated* by a high pressure in the kidneys [inanimate agent passive]

- (5) These sex differences (...) *can also be initiated* by injection of anterior pituitary extracts
[Janus-agent passive]
- (6) The invitation *hadn't been* very felicitously *phrased* [agentless passive]
- (7) I've always *been* far too *inclined* to treat important objects as part of my own petty
existence [attitudinal passive]
- (8) Gerald *was* suddenly very *annoyed* [emotive passive]
- (9) *Is* the thesis *finished*? [nonagentive passive/statal passive]
- (10) Cavill *was* *unimpressed* by this sally [compound passive]

This categorization relies on a large number of criteria. For example, adjectival characteristics include the possible modification of the *Ved* form by the intensifying adverb *very* (cf. sentence (8)) and the replacement of BE by copulas such as BECOME, GET or FEEL. Granger (1983) uses similar criteria in her corpus-based investigation of the passive in spoken registers. She also breaks down BE *Ved* constructions into several categories, but unlike Svartvik she reserves the term 'passive' to central passives 'which stand in direct alternation to a semantically equivalent active verbal group' (Granger, 1983: 108). In addition to this category, she distinguishes two categories of pseudo-passives (adjectival and verbal), peripheral and mixed combinations, and assigns a special status to BE *Ved* constructions that 'share all characteristics of passives but whose active counterpart is far less common', the 'usually passive' category (Granger, 1983: 112). Her aim in distinguishing this category was to encourage researchers not to limit themselves to lists of verbs that cannot passivize but to also highlight verbs 'that never or rarely activize' (Granger, 1983: 236).

The borders between these categories are characterized by a high degree of fuzziness and overlap, represented by Granger in the form of three interlocking circles (1983: 107). As pointed

out by Palmer (1987: 85), the criteria used are ‘matters of degree rather than absolute criteria’. There are ‘degrees of “adjectiveness”’ (Palmer, 1987: 89) and ‘numerous dubious cases where a corresponding active form might be thought possible’ (Palmer, 1987: 87). As rightly observed by Rowley-Jolivet (2001: 46), ‘[t]he boundary between adjectival and verbal passives has been the subject of much linguistic discussion, but with little consensus’.

The idea of a passive gradient is generally absent from ELT grammars. When the ambiguity of the BE *Ved* structure is mentioned at all, it is usually presented as a binary distinction between passives and adjectives which can easily be disambiguated in context (see, for example, Cowan, 2008: 403; Parrott, 2000: 292). Reference grammars do recognize the existence of a passive gradient but differ in the way they subcategorize the passive categories and the terms used to refer to them. Quirk *et al.* (1985: 167) put forward a ‘passive gradient’ made up of three categories, i.e. central passives, semi-passives and pseudo-passives, but only consider the first one as passive. Biber *et al.* (1999: 475) acknowledge that ‘[p]assive constructions form a fuzzy category’ which includes adjectival and stative (i.e. statal) passives, but decide to adopt ‘a relatively broad definition of passive, excluding only forms that are clearly adjectival in function’. This still leaves room for a great deal of uncertainty as there are no clear-cut criteria to establish adjectival status and, in the words of the authors themselves, the passive ‘forms a cline with the copular verb + participial adjective clause pattern’ (Biber *et al.*, 1999: 936). In another section of the grammar the authors note that they also exclude idioms such as *be bound to* and *be supposed to*, ‘which do not admit of any active-passive variation’ but include stative passives for which ‘it is difficult to draw a clear borderline’ (Biber *et al.*, 1999: 937). It is clear from this brief survey that the term ‘passive’ is highly ambiguous. In its strict, most usual sense, it refers to verbal constructions that stand in direct alternation with an active construction. However, it can also be used as an all-embracing term to refer to all BE *Ved* constructions

including those that display adjectival characteristics and have an indirect relation to the active or no active counterpart at all, or it can cover only specific subsets of these constructions. There is also great indeterminacy in the terminology used to refer to the structural categories, which adds to the confusion.

3 The Passive in L2 Studies

The passive has been the subject of many L2 studies. In these studies, the notion of passive is often taken for granted and examples show that the targeted forms are usually central passives. In corpus-based studies, very little information, if any, is given on the criteria used to select the BE *Ved* forms. Some of the studies, primarily those relying on experimental data, have focused on the grammaticality (or rather ungrammaticality) of passives with a view to understanding how the passive is processed by learners (Section 3.1). Studies based on learner corpus data, on the other hand, have investigated aspects of use: the frequency of the passive (Section 3.2) and its lexical and phraseological patterning (Section 3.3).

3.1 Misuse

A large number of L2 studies of the passive (e.g. Chung, 2014; Ju, 2000; Kondo, 2005; Oshita, 2000) focus on overpassivization of intransitive verbs like *happen*, *disappear* or *die*, resulting in erroneous sentences such as **John was died in 1968*. Most of these studies rely on experimental data, in particular controlled elicitation tasks and grammaticality judgments tests, and try to identify the factors that play a part in these errors, factors such as animacy, the presence or not of a transitive counterpart, L1 influence and proficiency level. While some factors prove to have a strong impact on the number and types of errors, several findings are inconclusive and even contradictory. This is the case, for example, of L1 effects, with Kondo (2005: 160) concluding that overpassivization is universal to second language learners and

Chung (2014) finding strong L1 effects. It is important to bear in mind, however, that most overpassivization studies are based on limited data from a small number of – usually Asian – learner populations. In a corpus study of overpassivization based on a much more diversified database, Granger (2013) found significant differences between learner populations, with some populations, particularly non-Asian ones, producing very few overpassivization errors or none at all.

3.2 Underuse

One aspect that cannot be investigated on the basis of experimental data is frequency of use. As pointed out by Hinkel (2004: 6) in her investigation of tense, aspect and voice, ‘when learners participate in controlled experiments that focus on forms of verbs, the requisite tasks are completed and verb forms are produced regardless of whether NNSs (non-native speakers) of English would actually venture to use these particular tenses, aspects and/or voice in real-life L2 text production’. Corpus data, however, give direct access to actual use and non-use. Hinkel (2004) compared academic texts produced by native speakers of English and learners from six L1 backgrounds (Arabic, Chinese, Indonesian, Japanese, Korean and Vietnamese) and found a significant underuse in NNS essays compared to native speaker texts. She added that this underuse is all the more significant as the native writers were novice writers who had not yet fully mastered the conventions of written academic discourse. Xiao (2007) noted a highly significant underuse in a study focused on Chinese learners. The same result was obtained for a wide range of learner populations (cf., for example, Granger, 1997, 2013). Although several authors attribute the underuse of the passive (at least partly) to a lower frequency of the passive in the learners’ mother tongue, more research is necessary to tease out developmental and L1-induced factors.

3.3 Lexical and phraseological patterning

The impact of lexical choices on the use (or not) of the passive is hardly ever taken into account in L2 studies. Yet, Möller (2017: 288), focusing on CLIL (Content and Language Integrated Learning) and non-CLIL learners of English, underlines the importance of such choices when noting that non-CLIL learners may give the impression that they ‘master the passive better than they actually do’ (Möller 2017: 290) because of the particularly high frequency of the verb ALLOW in the passive, which, she argues, is the result of the acquisition of the chunk *be allowed* rather than the acquisition of the passive as such.

The comparison of lexical bundles in native and non-native English has been the topic of several studies (see Ebeling and Hasselgård, Chapter 2 in this volume). Some of these phraseological patterns rely on the use of the passive. Allen (2010), for example, points out that lexical bundles made up of a passive verb and a prepositional phrase represent 6% of all lexical bundles in a corpus of research papers written in English by Japanese science students, as against 30% in native academic writing (Hyland, 2008). Interestingly, Allen (2010) also notes an overuse of certain passive lexical bundles, and in particular *it is (widely/well) known that* and *(it) can be said (that)*. The latter, he suggests, could have been transferred from the learners’ L1. Granger (1998) compares passive frames of the type ‘*it* + (modal) + passive verb (of saying/thinking) + *that*-clause’ (e.g. *it is said that...*; *it can be claimed that...*) with their active counterpart ‘*I* or *we/one/you* + (modal) + active verb (of saying/thinking) + *that*-clause’ (e.g. *I maintain that...*; *one could say that...*) in native and French-speaking learner corpus data. She shows that, while there is no significant difference in the use of the passive structure, the learners heavily overuse the active structure, with about four times as many occurrences in the learner corpus as in the native corpus. Shaw and Liu (1998), adopting a longitudinal perspective, reveal an increase in passive metadiscoursal phrases over time (e.g. *it can be concluded that*) and a corresponding

decrease in active metadiscoursal phrases (e.g. *as a conclusion I can say*). As appears from Salazar (2014: 118ff.), this tendency can even continue to the point where passive structures become overused and active structures underused.

4 The EFL-ESL Continuum

Our study of the passive will be carried out from an inter-varietal perspective, comparing native English with non-native English, but also comparing several non-native varieties with each other. Within non-native varieties of English, a distinction is traditionally drawn between EFL and ESL (see Quirk *et al.*, 1985: 4-5). The former describes a situation in which English is learned through instruction, with little or no exposure to the target language in everyday life, except for international purposes. The latter refers to a situation in which English is acquired through exposure in everyday life (e.g. via the national media), possibly combined with more formal instruction. The main criterion that is used to distinguish between EFL and ESL is the status of English in a country: if English is an official or semi-official language in the country, it is considered a second variety; otherwise, it is regarded as a foreign variety.

The traditional dichotomy between EFL and ESL has, however, given way to a continuum, recognizing the existence of a grey zone between typical EFL and typical ESL situations. Both sociolinguistic and linguistic considerations have led to this recognition. From a sociolinguistic point of view, it has become clear that some countries present a mixed EFL-ESL situation. Scandinavian countries, for example, are EFL countries in the sense that English does not have any official status, but they have come to resemble ESL countries because of the increasingly widespread use of English for intranational purposes (see Simensen, 2010). The continuum between EFL and ESL has also emerged linguistically, with some varieties presenting both EFL and ESL linguistic features. Edwards (2014), for example, demonstrates that the use of the

progressive in Dutch English displays characteristics of EFL (e.g. in terms of overall frequency) as well as characteristics of ESL (cf. extension to stative verbs). Multifactorial analyses have also shown that EFL and ESL varieties do not necessarily cluster in two separate groups. Rautonaho *et al.*'s (2018) analysis of the progressive, for instance, reveals the existence of clusters of EFL varieties (French and Polish English) and of ESL varieties (Nigerian and Singaporean English), but also one cluster bringing together Finnish English, an EFL variety, and Indian English, an ESL variety.

In this study, we compare typical EFL varieties (English produced by speakers of French, German, Korean and Serbian) with varieties that are claimed to display both EFL and ESL features: English produced by speakers of Chinese from Hong Kong, speakers of Dutch from the Netherlands, speakers of Norwegian and speakers of Tswana. Hong Kong English can be described as an ESL variety because of the colonial past of Hong Kong but is often claimed to have transitioned from ESL to EFL (Görlach, 2002: 109-110). English in the Netherlands is traditionally seen as an EFL variety 'moving towards ESL status' (Edwards, 2014: 175). Norway, as a Scandinavian country, has arguably experienced the same shift (see above). As for Tswana English, it appears to have a hybrid status, in between EFL and ESL (Gilquin & Granger, 2011: 75). For the sake of simplicity and despite their hybrid status, in what follows we refer to these varieties as ESL (or ESL-like) varieties.²

We are interested in comparing EFL and ESL learners to native speakers of English, with respect to the frequency of the passive, as well as its lexical preferences and phraseological sequences. In addition, we want to compare EFL and ESL learners with each other. Following

² It should be noted that we approach the EFL-ESL continuum from the starting point of an EFL corpus (see Gilquin, 2016: 228-235), rather than relying on the comparison of an EFL corpus and an ESL corpus (see Gilquin, 2016: 235-241), which means that we will not cover the full spectrum of the EFL-ESL continuum.

a usage-based view of language acquisition (see Diessel, 2014), we expect the passive to be used in a more native-like manner in ESL than in EFL. This is because the passive construction is assumed to gradually emerge, with all its (grammatical, discursive, stylistic, lexical and phraseological) features, as a result of exposure to many instances of the construction. ESL learners, with their enhanced exposure to the target language in different domains of everyday life, are therefore expected to have a better command of the passive construction than EFL learners, who receive limited input in the target language in more restricted contexts (mainly instructional ones). Given the marked, non-canonical nature of the passive construction (Huddleston & Pullum, 2002: 46), the degree of exposure should play a particularly important role in the production of the passive, as suggested by Ellis *et al.* (2015: 364): ‘Learning the usages that are normal or unmarked from those that are unnatural or marked requires a huge amount of immersion in the speech community’. Finally, since all learner populations are presumably situated differently along the EFL-ESL continuum, we can assume that there will also be some variation between the eight populations under study with respect to how native-like their use of the passive is. Our research questions can be summarized as follows:

- RQ1 Do learners of English use the passive in a native-like manner in terms of frequency?
- RQ2 Do learners of English use the passive in a native-like manner in terms of lexical preferences?
- RQ3 Do learners of English use the passive in a native-like manner in terms of phraseological sequences?
- RQ4 Do ESL learners use the passive differently from EFL learners (in terms of frequency, lexical preferences and phraseological sequences)?

RQ1 will involve the comparison of relative frequencies, while RQ2 will examine the passive ratio of individual verbs. RQ3 will rely on the qualitative analysis of some selected phraseological patterns in the passive. Finally, RQ4 will focus more precisely on the comparison between EFL and ESL with respect to all these aspects.

5 Data and Methodology

5.1 Data

The L2 data were taken from the third version of the International Corpus of Learner English (ICLE; Granger *et al.*, 2020), a corpus made up of (mainly) argumentative texts written by university students from different mother tongue backgrounds. Although ICLE is officially a corpus of EFL, some of its data were collected in more ESL-like contexts (see also Gilquin & Granger, 2011). The choice of ICLE samples for the present study relied on this specificity. Using the ICLE online interface, eight subcorpora were compiled: four representing EFL populations (French-speaking Belgian students [FR], German students [GE], Korean students [KR] and Serbian students [SE]) and four representing ESL-like populations (Dutch-speaking students from the Netherlands [DU], Chinese-speaking Hong Kong students [HK], Norwegian students [NO] and Tswana students [TS]). Two main criteria were applied to select the relevant data, namely students' native language and the institution where the data were collected. For the French subcorpus, for example, we kept only the texts written by students with French as a native language, and among those we excluded three texts that had been collected in institutions outside the French-speaking part of Belgium (one in Spain, one in Sweden and one in the Dutch-speaking part of Belgium). In total, 3,174 texts corresponding to 1,793,680 words were selected for analysis. As a native baseline, we used data from the Louvain Corpus of Native English Essays (LOCNESS), and more precisely the essays written by British and American university

students. Taken together, they constitute a corpus of 298 texts and 264,886 words. Table 4.1 shows the breakdown per corpus and, for ICLE, per variety and population. It should be noted that, because of the criteria that were applied, the subcorpora vary in size, from 386,041 words for the ICLE-HK subcorpus to only 134,391 words for the ICLE-DU subcorpus.

Table 4.1 Breakdown of the corpora, by number of texts and word count

	Number of texts	Number of words
ICLE	3,174	1,793,680
EFL	1,467	860,314
FR	311	203,326
GE	431	229,712
KR	400	224,655
SE	325	202,621
ESL	1,707	933,366
DU	92	134,391
HK	780	386,041
NO	316	212,324
TS	519	200,610
LOCNESS	298	264,886

5.2 Methodology

Our search for passive forms centred around 20 verbs that are often used in the passive voice in expert academic writing. In order to choose these verbs, we searched a part-of-speech (POS) tagged version of the academic component of the BNC Baby, made up of c. 1 million words. All lexical verbs tagged as past participles and immediately following the lemma BE were extracted, together with their frequency in this passive pattern. The 15 most frequent forms were selected for analysis: *considered, described, found, given, known, made, needed, provided,*

said, seen, shown, suggested, taken, used and *written*. In line with the concept of passive gradient (Section 2.3), we also aimed to investigate passive forms that are usually not taken into consideration in studies of the passive. With this aim in view, we looked for forms ending in *-ed* and tagged as adjectives, immediately preceded by the lemma BE. We selected the 5 most frequent forms in the corpus, namely *concerned, interested, involved, satisfied* and *supposed*. It is interesting to note that the same forms used in similar contexts were sometimes tagged as past participles and sometimes as adjectives. Compare (11), where *given* is tagged as a past participle (VVN), with (12), where it is tagged as an adjective (JJ), despite its clearly verbal nature. This shows that the tagging of forms as either past participles or adjectives in *-ed* is arbitrary at times, which further justifies our decision to also count as passives forms that are tagged as adjectives.

(11) While reading a piece of literature, the audience *is given_VVN* more time to accept the knowledge with which it has been confronted. (LOCNESS)

(12) Because the knowledge is conveyed in a familiar language, through a familiar medium of text, the reader *is given_JJ* time in which to accept his own misconceptions (...) (LOCNESS)

On the basis of a POS tagged version of LOCNESS and ICLE, we extracted all the occurrences of the 20 selected forms preceded by the lemma BE, allowing for the possibility of one word between these two elements.³ All the concordance lines were examined manually to exclude

³ The optimal distance between BE and the *Ved* form was determined by means of a pilot study which compared the number of passive forms retrieved and the amount of noise (i.e. irrelevant hits) in ICLE-FR and ICLE-DU when varying the number of intervening words, from 0 to 3. The results were very similar in both subcorpora, with three-quarters of the data having no intervening words. By allowing for one word between BE and the *Ved* form,

cases that were clearly not passive constructions, e.g. *It is a **given** that school uniform will turn into the mainstream medium for connecting to the old memories*. Phrasal verbs (e.g. *Recycling will **be given** up*) were also discarded on the grounds that they represent different verbs and tend to occur in different phraseological patterns. All the other concordances (765 in LOCNESS and 3,968 in ICLE) were kept for analysis, even cases where the form was more adjectival (e.g. *However, only few managers **are concerned** about the harmful effects of their business policies*), with the exception of compound adjectives spelled with a hyphen (e.g. *well-known*). This was done to avoid any arbitrary distinction along the passive gradient – one that, moreover, would most probably not correspond to any actual distinction in the learner’s mind. All these forms will be referred to in the analysis as passives, with no distinction between central and more peripheral cases.⁴

For each verb in each (sub)corpus, we calculated the relative frequency of the passive (number of occurrences of the passive out of the total number of words in the (sub)corpus). We also computed the passive ratio, which represents the proportion of passive uses out of all the occurrences of the lemma. The passive ratio is normally computed by looking at verbal forms only (see Section 2.2). However, since we also included some forms tagged as adjectives in our analysis, we computed the passive ratio by adding up the verbal and adjectival forms. For

it was possible to retrieve 15% more of the data, with very little noise. While allowing for 2-3 intervening words would retrieve 10% more of the data, it would also result in a great number of irrelevant hits, which would have to be discarded manually. Taking one intervening word into account thus seemed to be a good compromise in terms of precision and recall.

⁴ It must be emphasized that some passive constructions will not have been retrieved by our procedure: constructions with the GET passive auxiliary, constructions with more than one word between the passive auxiliary and the passive form, passive constructions with no passive auxiliary (e.g. *As argued by...*), with mistagged elements (e.g. a past participle tagged as a past tense) or with non-standard forms (e.g. BE *know*, BE *gaven*).

involved in LOCNESS, for example, we extracted 29 passive forms, of which 10 were tagged as past participles and 19 as adjectives. When searching for the frequency of the lemma INVOLVE, we took account of both INVOLVE as a verb (93 occurrences) and INVOLVED as an adjective (41 occurrences), for a total of 134 occurrences. The passive ratio for INVOLVE in LOCNESS was thus calculated by dividing the sum of the passive forms (29) by the sum of the (verbal and adjectival) lemma (134), which resulted in a passive ratio of 21.6%. For both frequencies and passive ratios, statistical significance was measured through a chi-square test with Bonferroni correction.

6 Passive Frequency

Overall, grouping together the 20 verbs under study, the passive appears to be underused by the learners, with a relative frequency of 221.2 per 100,000 words in our ICLE sample, to be compared with 288.8 in LOCNESS. This underuse is found to characterize both EFL and ESL populations, despite a slightly higher (but not significantly different) frequency in the ESL data (226.2 per 100,000 words) than in the EFL data (215.9 per 100,000 words).

Table 4.2 shows the absolute and relative frequency of the passive in LOCNESS and the different ICLE subcorpora, with the asterisk indicating a statistically significant difference between the ICLE subcorpus and LOCNESS. The highest frequency is found in ICLE-DU and ICLE-FR, which include proportionally more passive verbs than LOCNESS, although the difference is not statistically significant. All the other ICLE subcorpora include proportionally fewer passive verbs than LOCNESS, and all of them, except ICLE-TS, present a significant underuse of the passive in comparison with the native baseline (as signalled by the minus sign).

Table 4.2 Absolute and relative frequency (per 100,000 words) of the passive in LOCNESS and ICLE subcorpora

Corpus	Absolute frequency	Relative frequency
ICLE-DU	417	310.3
ICLE-FR	616	303.0
LOCNESS	765	288.8
ICLE-TS	555	276.7
ICLE-SE* ⁻	472	232.9
ICLE-NO* ⁻	483	227.5
ICLE-GE* ⁻	393	171.1
ICLE-HK* ⁻	656	169.9
ICLE-KR* ⁻	376	167.4

Note: the asterisk indicates a statistically significant difference between the ICLE subcorpus and LOCNESS; the minus sign indicates a significant underuse.

These results confirm earlier studies that have brought to light a tendency, among learners, to underuse the passive voice (Section 3.2), but they also show that not all learner populations are (equally) affected by this underuse. Interestingly, no clear distinction emerges between the EFL and ESL-like populations: not only is there no statistically significant difference in frequency between the EFL and ESL data (see above), but if we look at the populations with the highest and lowest frequency of passives (top and bottom of Table 4.2), we find both types of learners represented (ICLE-DU/ICLE-FR and ICLE-KR/ICLE-HK). This suggests that, contrary to what a usage-based view of language acquisition would predict, differences in the degree of exposure to the target language do not necessarily lead to differences in the frequency of the passive.

While most studies of the passive merely consider its overall frequency, we were particularly interested in possible variation between the different verbs, as a way of measuring the impact of lexis on the use of the passive. The results for the relative frequency revealed striking disparities between certain verbs. In ICLE-TS, for example, which shows no overall difference with LOCNESS, four verbs turn out to be significantly underused in the passive (MAKE, SEE, SHOW, WRITE), whereas three verbs are significantly overused (TAKE, GIVE, PROVIDE). In ICLE-FR, another subcorpus with no significant difference in comparison with LOCNESS, three verbs are significantly underused in the passive (USE, SEE, SHOW) and four are significantly overused (FIND, SAY, CONCERN, INTEREST). While two of the overused passive forms in ICLE-FR – CONCERN and INTEREST – are mostly tagged as adjectives in the corpora, it should be emphasized that some of their uses are clearly verbal (e.g. *as far as x BE concerned*). Thus, the overuse of these forms does not simply reflect an overuse of adjectives. These differences in the relative frequency of certain passive verbs demonstrate that the passive is not necessarily a grammatical rule that applies across the board, affecting all transitive verbs in the same way. In order to further investigate the lexical aspects of the passive, in the next section we rely on the measure of passive ratio.

7 The Lexis-Grammar Interface

7.1 The passive ratio

The relative frequency of a verb in the passive could be influenced by the overall frequency of this verb. A passive verb could thus be underused in a learner corpus simply because the verb itself is less often used by the learners (e.g. due to the learners' unfamiliarity with this verb or as a consequence of the topic on which the learners had to write). The passive ratio provides a more reliable measure by taking the overall frequency of the verb into account

and by calculating the proportion of passive uses out of all the occurrences (active and passive) of the verb.

The first column of Table 4.3 shows the passive ratio, in descending order, of the 20 verbs under study in LOCNESS. The passive ratios differ widely, from 73.9% for INTEREST to 6.1% for SAY. Interestingly, the verbs with a high passive ratio do not only include forms that were selected via a search for past participles (e.g. CONSIDER, SEE, USE), but also, and even more prominently, forms that were selected via a search for adjectives in *-ed*. This is the case of the top three passives, INTEREST, SUPPOSE and CONCERN; the other two forms chosen among adjectives, namely SATISFY and INVOLVE, are well ranked too, with a passive ratio higher than 20%. Among these forms tagged as adjectives, some have clear adjectival features in most of their occurrences (cf. INTEREST, SATISFY), while others are actually mostly used as verbal forms (cf. SUPPOSE, CONCERN), and yet others combine adjectival and verbal features. This justifies our inclusive approach, which consisted in taking both past participles and adjectives in *-ed* into account. This also suggests that the so-called peripheral representatives of the passive may in fact be more central than usually assumed, at least quantitatively speaking.

The comparison of these results with those in the second column of Table 4.3 makes it possible to highlight differences in passive ratio between native and non-native English (the boldface and the asterisks indicate statistically significant differences and are accompanied by a plus sign for a ratio significantly higher in ICLE than in LOCNESS and a minus sign for a significantly lower ratio). Despite a difference in ranking, the top three forms are the same in LOCNESS and ICLE, and their passive ratios do not differ significantly. Further down the list, the rankings present many more differences, and several verbs in ICLE have a passive ratio that is significantly different from that in LOCNESS. In most cases, this difference corresponds to a

Table 4.3 Passive ratio of the 20 verbs in LOCNESS and ICLE (in descending order)

LOCNESS passive ratio		ICLE passive ratio	
(%)		(%)	
INTEREST	73.9	SUPPOSE	67.9
SUPPOSE	58.1	INTEREST	65.7
CONCERN	52.9	CONCERN	53.3
CONSIDER	46.2	DESCRIBE	33.7* ⁺
SATISFY	37.5	SATISFY	33.3
SEE	24.1	CONSIDER	30.9* ⁻
USE	22.3	INVOLVE	19.6
INVOLVE	21.6	GIVE	12.9
GIVE	16.5	WRITE	12.5
NEED	15.7	FIND	12.1
SHOW	14.8	USE	11.4* ⁻
DESCRIBE	13.7	SEE	10.4* ⁻
WRITE	12.4	SAY	8.5
MAKE	10.2	NEED	7.9* ⁻
FIND	10.0	SUGGEST	7.6
SUGGEST	9.5	TAKE	7.5
TAKE	8.2	MAKE	7.1* ⁻
PROVIDE	7.1	PROVIDE	7.1
KNOW	6.9	SHOW	6.2* ⁻
SAY	6.1	KNOW	5.0

Note: the boldface and the asterisks indicate statistically significant differences and are accompanied by a plus sign for a ratio significantly higher in ICLE than in LOCNESS and a minus sign for a significantly lower ratio.

lower passive ratio in ICLE (CONSIDER, USE, SEE, NEED, MAKE, SHOW); in only one case does it correspond to a higher passive ratio in ICLE (DESCRIBE). The difference in passive ratio

between LOCNESS and ICLE is sometimes quite considerable, with some passive ratios being at least twice as high in one corpus as in the other (cf. DESCRIBE, USE, SEE, NEED, SHOW).

After considering ICLE as a whole, we made a distinction between the ICLE subcorpora corresponding to EFL varieties and those corresponding to ESL-like varieties. Figure 4.1 represents the passive ratio of the different verbs in EFL and ESL, with the line showing the passive ratio in LOCNESS. Few bars in the graph extend far beyond the native line. Overall (i.e. all verbs taken together), the passive ratios in both ICLE-EFL (12%) and ICLE-ESL (11.2%) are significantly lower than the passive ratio in LOCNESS (15.6%). DESCRIBE is the only verb whose passive ratio is significantly higher in non-native writing (EFL) than in native writing. While EFL and ESL follow generally similar tendencies, certain verbs display significant differences between EFL and ESL: WRITE, INVOLVE, USE, CONSIDER and CONCERN have significantly higher passive ratios in EFL than in ESL, whereas TAKE, SUGGEST and GIVE have significantly higher passive ratios in ESL than in EFL. It is sometimes EFL that is closer to the native baseline (cf., e.g., FIND), and sometimes it is ESL (cf., e.g., GIVE). It is thus not exactly the case that ESL learners, with their higher exposure to English, display a more native-like behaviour than EFL learners. The picture turns out to be much more nuanced.

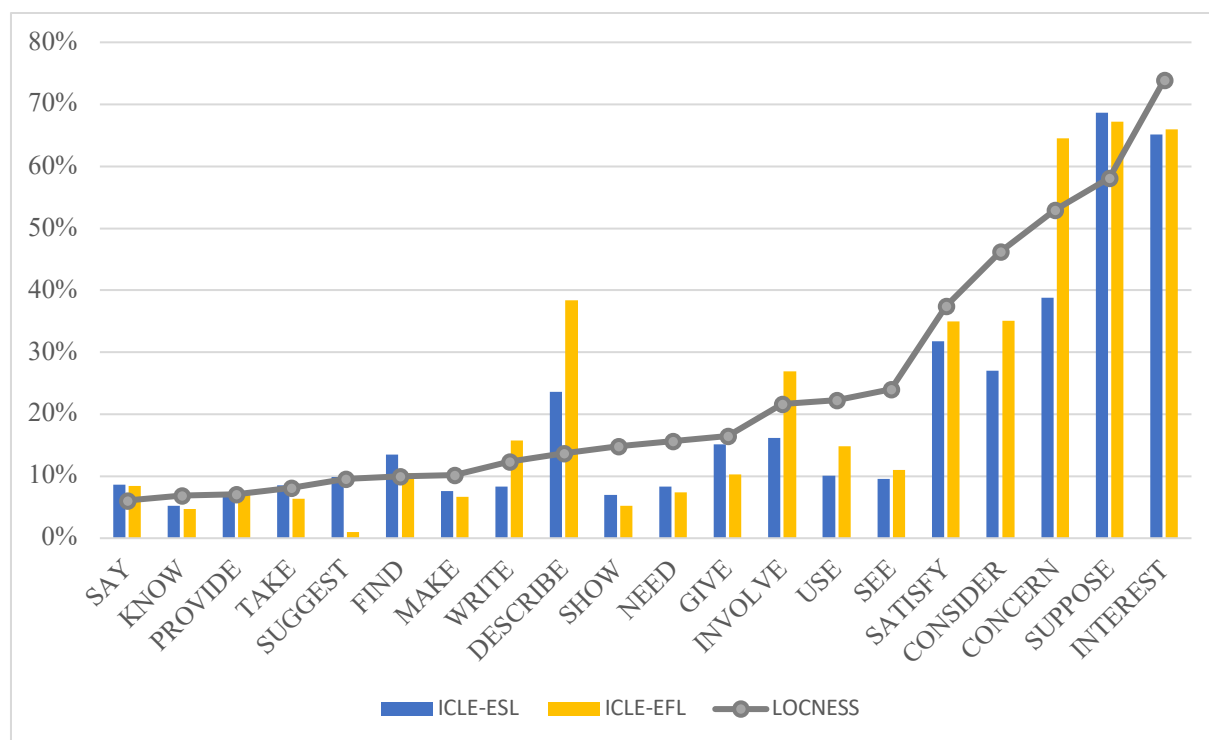


Figure 4.1 Passive ratio per verb in LOCNESS, ICLE-ESL and ICLE-EFL

Turning now to the specific learner populations, it appears that, besides variation between verbs, there is also inter-varietal variation. Table 4.4 summarizes the results of statistical significance tests comparing, for each verb, the passive ratios in LOCNESS and the individual ICLE subcorpora. A majority of the cells (121 out of 160, i.e. three-quarters) are empty, indicating that there is no statistically significant difference between LOCNESS and the ICLE subcorpus. In 30 cases, a minus sign reveals a significantly lower passive ratio in the ICLE subcorpus, while in 9 cases, a plus sign reveals a significantly higher passive ratio in the ICLE subcorpus. Of the 20 verbs, 5 display no significant differences at all with native English across the ICLE subcorpora, namely INTEREST, PROVIDE, SATISFY, SUGGEST and WRITE. Five other verbs present a significant difference for one subcorpus only: CONCERN in ICLE-FR (+), DESCRIBE in ICLE-SE (+), FIND in ICLE-HK (+), INVOLVE in ICLE-HK (–) and KNOW in ICLE-NO (–). In one case, for the verb SEE, all the ICLE subcorpora present a significantly lower passive ratio in

comparison with the native baseline. In the other cases, we see a mixture of non-significant, significantly lower and/or significantly higher results.

Table 4.4 Comparison of the passive ratios in the ICLE subcorpora with the native baseline

Verb	ICLE-EFL				ICLE-ESL			
	FR	GE	KR	SE	DU	HK	NO	TS
CONCERN	+							
CONSIDER		–	–			–		
DESCRIBE				+				
FIND						+		
GIVE			–			–		+
INTEREST								
INVOLVE						–		
KNOW							–	
MAKE			–	–		–		
NEED		–		–			–	–
PROVIDE								
SATISFY								
SAY	+				+			
SEE	–	–	–	–	–	–	–	–
SHOW	–	–				–		
SUGGEST								
SUPPOSE			–	+				
TAKE			–			–		+
USE				–	+	–		
WRITE								

On the basis of a list of the exact passive ratios for each verb in LOCNESS (see Table 4.3) and in the different ICLE subcorpora (see Appendix, Table A.1), a hierarchical cluster analysis was carried out, which made it possible to produce the dendrogram shown in Figure 4.2.⁵ The dendrogram provides a general overview of the overall degree of similarity between the different populations. Three significantly distinct clusters emerge from the analysis: one with ICLE-KR and ICLE-HK, another one with ICLE-TS and ICLE-SE, and a third one with the remaining ICLE subcorpora – FR, DU, NO, GE – and the native corpus. Interestingly, all three clusters include at least one EFL variety and one ESL-like variety. This seems to challenge the usage-based prediction that different types and amounts of exposure to the target language should lead to different linguistic behaviours in the L2, but it confirms the results of other multifactorial analyses showing that EFL and ESL do not necessarily cluster in two separate groups (see Section 4). As for closeness to native English, in addition to ICLE-DU and ICLE-NO – two ESL-like varieties –, ICLE-GE and ICLE-FR – two EFL varieties – turn out to cluster with LOCNESS and hence to resemble the native variety most closely. It will also be noticed that this cluster includes three Germanic L1s (Dutch, Norwegian and German). This could be due to the typological closeness of these languages to English, to the language proficiency of these learner populations,⁶ or to other factors which cannot be further investigated here but which could outweigh the distinction between EFL and ESL.

⁵ The hierarchical cluster analysis used the Euclidean distance and the Ward agglomerative method. In the dendrogram, AU stands for ‘approximately unbiased’ and BP for ‘bootstrap probability’. The identification of the significant clusters relied on the AU values. We thank Samantha Laporte for her help with this analysis.

⁶ The rating of a random sample of twenty essays per subcorpus according to the Common European Framework of Reference for Languages (see Granger *et al.*, 2020: 12) shows that a majority of the essays in ICLE-DU, ICLE-GE and ICLE-NO are rated as either C1 or C2; the same is true of ICLE-FR. This is to be contrasted with ICLE-KR, ICLE-TS and ICLE-HK, which all have a majority of B2 or lower essays. ICLE-SE is an exception in this respect, since it includes a majority of C1/C2 essays but does not cluster with LOCNESS in the dendrogram.

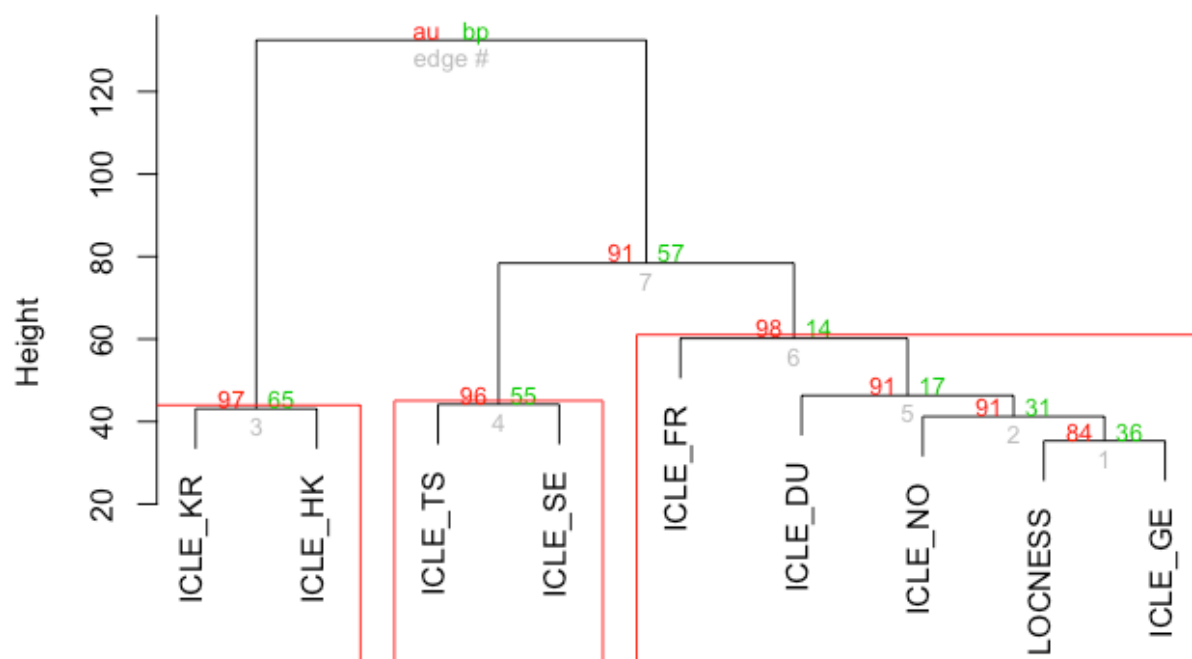


Figure 4.2 Dendrogram of the hierarchical cluster analysis of passive ratios

Note: AU stands for ‘approximately unbiased’; it corresponds to the values on the left. BP stands for ‘bootstrap probability’; it corresponds to the values on the right.

7.2 Phraseological sequences

Passive verbs have their own phraseologies. For example, they have their preferred collocates (e.g. *an action/a position/a step is taken*) and form (semi-)idiomatic expressions (e.g. *is taken into account/into consideration/for granted*). However, these types of phrasemes are not specific to passive use and do not feature in sufficient numbers with the selected verbs in our corpora. They will therefore not be investigated here. Instead, we investigate two types of patterns that form typical chunk-like passive sequences and are sufficiently frequent to allow for comparisons between the subcorpora. We first focus on passive patterns with the verb *CONCERN*, one of the verbs that stood out as being significantly overused in only one subcorpus, i.e. ICLE-FR. Our objective is to assess to what extent the overuse of the verb might be explained by its phraseological patterning. We then investigate impersonal *it*-patterns that occur with several of the selected verbs. The analysis centres on the verbs *KNOW* and *SAY*, which

contain sufficient instances of the pattern to allow for meaningful analysis, although in view of the size of the subcorpora, the figures remain quite low. The approach adopted in this section involves a close examination of concordance lines and is therefore essentially qualitative.

7.2.1 BE concerned patterns

The two main phraseological patterns of BE *concerned* in LOCNESS are BE *concerned with* and *as far as x BE concerned*. Together they account for 78% of the occurrences of BE *concerned* in the corpus. The ICLE learners are familiar with these patterns and display relatively similar proportions of use (between 52% and 73%, depending on the subcorpus), except for two populations that clearly stand out. ICLE-HK displays a very low proportion of 11%. This is mainly due to the fact that the Hong Kong learners tend to use BE *concerned* on its own, as illustrated by examples (13) and (14). This finding shows that passive ratio is not all: the Hong Kong learners' passive ratio for CONCERN does not differ quantitatively from the native baseline, but the qualitative analysis reveals that many of the forms are unidiomatic. ICLE-FR, on the other hand, displays a very high proportion of phraseological uses (91%), mainly due to the frequent use of the *as far as x BE concerned* pattern, which alone accounts for 80% of the passive uses (vs only 30% in LOCNESS). In the majority of cases (56 out of 90 occurrences) the French-speaking learners use the pattern at the beginning of the sentence as a topic introducer (examples (15) and (16)). Two factors could account for this overuse: form-function mapping from the frequent – albeit active – formula *en ce qui concerne x* in the learners' L1, reinforced by a general tendency among learners to cling on to a limited number of formulae that they feel comfortable with, the so-called 'teddy bear' effect (Hasselgård, 2019; Hasselgren, 1994). It is the overuse of this pattern that accounts for the significantly higher passive ratio of CONCERN in ICLE-FR. If the pattern is excluded, the passive ratio decreases from 82% to 47% and it becomes non-significantly different from LOCNESS. This example

therefore highlights the importance of taking both lexical and phraseological factors into account in analyses of the passive.

- (13) However, the cost of the railway is one of the matter *being* most *concerned*. (ICLE-HK)
- (14) At the same time, the production of importation of cigarette should also *be concerned*. (ICLE-HK)
- (15) *As far as* Europe *is concerned*, what we can observe is a marked tendency to specialization (ICLE-FR)
- (16) *As far as* the garden *is concerned*, it is divided into two parts (ICLE-FR)

7.2.2 Impersonal *it*-patterns

The *it*-patterns with KNOW and SAY are instantiated in all the ICLE subcorpora.⁷ The proportion of passive uses they represent varies greatly but does not reveal a distinct EFL-ESL divide. The patterns with KNOW are especially common in ICLE-HK and ICLE-SE, where they account for more than half of the passive verb forms (54% and 75% respectively, compared to 8% in LOCNESS). ICLE-HK displays a similar overuse for the pattern with SAY (77%) and also displays more impersonal patterns with SEE and SUGGEST than the other groups. This overuse may be due to teaching. Milton and Freeman (1996: 130) observe that Hong Kong learners are ‘taught indiscriminately to mark their discourse with a subset of lexical expressions which are used sparingly, if at all, by native speakers’, a practice that can reinforce learners’ general tendency to overuse a small set of ‘phraseological teddy bears’ (Hasselgård, 2019). A similar tendency to overuse the *it can be said* pattern was noted by Allen (2010) for Japanese learners and attributed to L1 transfer (see Section 3.3). Although L1 transfer does not account

⁷ With one exception: there is no occurrence of the *it*-pattern with KNOW in ICLE-NO.

for the use of this passive pattern by French-speaking learners (French does not have a direct equivalent of *it can be said*), it is clearly a strong potential factor in the use of the corresponding active structure. While ICLE-FR only contains 12 occurrences of the passive pattern with SAY, it contains 125 occurrences of the active *we/you/one* (aux.) *say that* pattern, the most frequently used sequence being *we can/could say that*, a literal translation of the French discourse formula *on peut/pourrait dire que*. This tendency is not shared by the Hong Kong learners, who rarely use the active pattern. This example illustrates the competition that can exist between an active and a passive phraseological pattern, the winner in the case of the French-speaking learners being the active pattern, presumably because of the influence of the corresponding French pattern.

Over and above the number of *it*-patterns, it is interesting to look at their different lexicalizations. Each pattern allows for a great deal of variation but this variation is not infinite: the patterns are not fully prefabricated but keep a degree of prefabrication. Some of the patterns used by the ICLE learners are fully idiomatic (see examples (17) and (18)), others are much less idiomatic, as illustrated by examples (19) to (22). Such instances tend to show that learners are not yet fully aware of the phraseological nature of these patterns. Their use can thus be described as midway between Sinclair's (1991) 'open-choice principle' and 'idiom principle'.

- (17) *it is said* that recycling is an expensive method (ICLE-HK)
- (18) *It can be said* that reading is more suitable for people who prefer spending time alone than socializing with others (ICLE-SE)
- (19) *It is not said without reason* that once you have a television you got the whole world (...) (ICLE-NO)

- (20) *It can be well known* that the target of study is to gain knowledge in the school (ICLE-HK)
- (21) *It is wide known* that university degree is the first and basic step in the life of the future (ICLE-SE)
- (22) *It has also been clearly seen and studied* that poverty can also cause HIV/AIDS epidemic in Africa (ICLE-TS)

8 Conclusion

This chapter has adopted a lexico-grammatical perspective to study the use of the passive (including its less central representatives) by EFL and ESL-like learners. It has revealed an overall underuse of the passive in EFL and ESL, visible among Serbian, Norwegian, German, Hong Kong and Korean learners (RQ1). It has also demonstrated that different verbs display different degrees of attraction towards the passive and that learners do not necessarily have the same lexical preferences as native writers (RQ2). Some verbs (e.g. PROVIDE) are closer to the native baseline than others (e.g. SEE) and some populations (e.g. ICLE-DU) display more native-like lexical preferences than others (e.g. ICLE-HK). In terms of phraseological sequences (RQ3), certain distinctive patterns emerge from the learner corpus data (e.g. *as far as x BE concerned* in ICLE-FR), among which some unidiomatic phrases that seem to be the result of both the ‘open-choice principle’ and the ‘idiom principle’. As regards the comparison between EFL and ESL (RQ4), overall they appear to display similar tendencies, most notably the general underuse of the passive. While a few differences have been highlighted (e.g. in the passive ratio of certain individual verbs), these do not necessarily go in the direction predicted by a usage-based perspective, i.e. with the ESL learners using the passive in a more native-like manner than the EFL learners. Several explanations could be put forward to explain these unexpected results. First, they could be explained by the usage-based view itself: ESL learners

receive more input, but this input may not be native-like in the first place; it is also likely to be mostly spoken or spoken-like, with relatively few instances of the passive, which is more typical of writing. This underlines the importance of considering the nature of the input in a usage-based perspective. Our results could also be explained by the impact of other factors such as L1 influence, proficiency level, teaching and developmental effects, or date of compilation of the (sub)corpora, which may outweigh the impact of the EFL-ESL continuum. Another possible explanation might be that the ESL-like populations that we have selected display too many traces of EFL and therefore do not differ fundamentally, or in a systematic way, from the EFL populations examined. Our study, however, does not make it possible to test the validity of these explanations. Replication studies are needed to assess the generalizability of our findings, based on different L2 corpora and/or relying on expert rather than novice baseline data, and taking more factors into account.

The study has also brought to light the very important role that lexico-grammar plays in passive constructions, both through the widespread use of forms tagged as adjectives (some of which are actually fully verbal or combine adjectival and verbal features) and through the attraction of certain verbs to the passive voice. The BE passive has traditionally been seen as a purely grammatical phenomenon, unlike the GET passive, whose lexical aspects are often recognized in the literature. Rühlemann (2007: 121), for example, writes the following about the GET passive: ‘To be sure, it *is* a grammatical structure, but it is at least equally, if not more so, a lexical structure’ (emphasis original). The same, it could be argued, applies to the BE passive. In terms of pedagogical implications, this means that learners should not only (or not mainly) be taught how to construct a passive structure but, as Salazar (2014: 127) puts it, they ‘should learn which structures should be used with which words and in which contexts’. More effort should therefore be channelled into the teaching of phraseological aspects of the passive, and

the development of tools and methods that can facilitate the acquisition of such aspects. The passive is clearly not the only linguistic structure that could benefit from a lexico-grammatical approach to grammar, but it is undoubtedly a strong candidate to start the ball rolling.

References

- Allen, D. (2010) Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education* 1, 105-127.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. Harlow, UK: Pearson.
- Bunting, J. D. and Diniz, L. (with Reppen, R.) (2013) *Grammar and Beyond 4*. Cambridge: Cambridge University Press.
- Chung, T. (2014) Multiple factors in the L2 acquisition of English unaccusative verbs. *IRAL* 52 (1), 59-87.
- Conrad, S. and Biber, D. (2009) *Real Grammar. A Corpus-Based Approach to English*. New York: Pearson.
- Cowan, R. (2008) *The Teacher's Grammar of English*. Cambridge: Cambridge University Press.
- Diessel, H. (2014) Usage-based linguistics. In M. Aronoff (ed.) *Oxford Bibliographies in 'Linguistics'*. New York: Oxford University Press.
- Edwards, A. (2014) The progressive aspect in the Netherlands and the ESL/EFL continuum. *World Englishes* 33 (2), 173-194.
- Ellis, N. C., Simpson-Vlach, R., Römer, U., Brook O'Donnell, M. and Wulff, S. (2015) Learner corpora and formulaic language in second language acquisition research. In S. Granger,

- G. Gilquin and F. Meunier (eds) *The Cambridge Handbook of Learner Corpus Research* (pp. 357-378). Cambridge: Cambridge University Press.
- Garner, B. A. (2016) *Garner's Modern English Usage. Fourth edition*. Oxford: Oxford University Press.
- Gilquin, G. (2016) Discourse markers in L2 English: From classroom to naturalistic input. In O. Timofeeva, A.-C. Gardner, A. Honkapohja and S. Chevalier (eds) *New Approaches to English Linguistics: Building Bridges* (pp. 213-249). Amsterdam: Benjamins.
- Gilquin, G. and Granger, S. (2011) From EFL to ESL: Evidence from the International Corpus of Learner English. In J. Mukherjee and M. Hundt (eds) *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap* (pp. 55-78). Amsterdam: Benjamins.
- Görlach, M. (2002) English in Singapore, Malaysia, Hong Kong, Indonesia, the Philippines ... a second or a foreign language? In M. Görlach (ed.) *Still More Englishes* (pp. 99-117). Amsterdam: Benjamins.
- Granger, S. (1983) *The Be + Past Participle Construction in Spoken English with Special Emphasis on the Passive*. Amsterdam: Elsevier Science Publishers.
- Granger, S. (1997) Automated retrieval of passives from native and learner corpora: Precision and recall. *Journal of English Linguistics* 25 (4), 365-374.
- Granger, S. (1998) Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.P. Cowie (ed.) *Phraseology: Theory, Analysis, and Applications* (pp. 145-160). Oxford: Oxford University Press.

- Granger, S. (2013) The passive in learner English: Corpus insights and implications for pedagogical grammar. In S. Ishikawa (ed.) *Learner Corpus Studies in Asia and the World* (pp. 5-15). Kobe: Kobe University, School of Languages and Communication.
- Granger, S., Dupont, M., Meunier, F., Naets, H. and Paquot, M. (2020) *The International Corpus of Learner English, Version 3*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Gries, S.Th. and Stefanowitsch, A. (2004) Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9 (1), 97-129.
- Hasselgård, H. (2019) Phraseological teddy bears: Frequent lexical bundles in academic writing by Norwegian learners and native speakers of English. In M. Mahlberg and V. Wiegand (eds) *Corpus Linguistics, Context and Culture* (pp. 339-362). Berlin: De Gruyter.
- 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 (2), 237-260.
- Hinkel, E. (2002) Why English passive is difficult to teach (and learn). In E. Hinkel and S. Fotos (eds) *New Perspectives on Grammar Teaching in Second Language Classrooms* (pp. 233-260). Mahwah, NJ: Lawrence Erlbaum.
- Hinkel, E. (2004) Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research* 8 (1), 5-29.
- Huddleston, R. and Pullum, G. K. (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

- Huddleston, R. and Pullum, G. K. (2005) *A Student's Introduction to English Grammar*. Cambridge: Cambridge University Press.
- Hyland, K. (2008) As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27 (1), 4-21.
- Ju, M. K. (2000) Overpassivization errors by second language learners: The effect of conceptualizable agents in discourse. *Studies in Second Language Acquisition* 22 (1), 85-111.
- Kondo, T. (2005) Overpassivization in second language acquisition. *IRAL* 43 (2), 129-161.
- Larsen-Freeman, D. and Celce-Murcia, M. (with Frodesen, J., White, B. and Williams, H.) (2016) *The Grammar Book: Form, Meaning, and Use for English Language Teachers. Third edition*. Boston, MA: National Geographic Learning.
- Lehmann, H. M. and Schneider, G. (2009) Parser-based analysis of syntax-lexis interactions. In A. H. Jucker, D. Schreier and M. Hundt (eds) *Corpora: Pragmatics and Discourse* (pp. 477-502). Amsterdam: Rodopi.
- Milton, J. and Freeman, R. (1996) Lexical variation in the writing of Chinese learners of English. In C. Percy, C. Meyer and I. Lancashire (eds) *Synchronic Corpus Linguistics* (pp. 121-131). Amsterdam: Rodopi.
- Möller, V. (2017) *Language Acquisition in CLIL and non-CLIL Settings: Learner Corpus and Experimental Evidence on Passive Constructions*. Amsterdam: Benjamins.
- Oshita, H. (2000) *What is happened* may not be what appears to be happening: A corpus study of 'passive' unaccusatives in L2 English. *Second Language Research* 16 (4), 293-324.
- Palmer, F. R. (1987) *The English Verb*. Longman Linguistics Library. Harlow, UK: Addison Wesley Longman.

- Parrott, M. (2000) *Grammar for English Language Teachers*. Cambridge: Cambridge University Press.
- Pullum, G. (2014) Fear and loathing of the English passive. *Language and Communication* 37, 60-74.
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- Rautioaho, P., Deshors, S. C. and Meriläinen, L. (2018) Revisiting the ENL-ESL-EFL continuum: A multifactorial approach to grammatical aspect in spoken Englishes. *ICAME Journal* 42, 41-78.
- Rowley-Jolivet, E. (2001) Activating the passive: A comparative study of the passive in scientific conference presentations and research articles. *Recherche et pratiques pédagogiques en langues de spécialité. Les Cahiers de l'Apliut* 20 (4), 38-52.
<http://journals.openedition.org/apliut/4750>
- Rühlemann, C. (2007) Lexical grammar: The GET-passive as a case in point. *ICAME Journal* 31, 111-127.
- Salazar, D. (2014) *Lexical Bundles in Native and Non-native Scientific Writing: Applying a Corpus-based Study to Language Teaching*. Amsterdam: Benjamins.
- Shaw, P. and Liu, E. T. K (1998) What develops in the development of second-language writing? *Applied Linguistics* 19 (2), 225-254.
- Simensen, A. M. (2010) English in Scandinavia: A success story. In D. Wyse, R. Andrews and J. Hoffman (eds) *The Routledge International Handbook of English, Language and Literacy Teaching* (pp. 472-483). New York, NY: Routledge.
- Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- Svartvik, J. (1966) *On Voice in the English Verb*. The Hague: Mouton.
- Swales, J. M. and Feak, C. B. (2004) *Academic Writing for Graduate Students: Essential Tasks and Skills. Second edition*. Ann Arbor, MI: University of Michigan Press.
- Vince, M. (2008) *Macmillan English Grammar in Context. Advanced*. Oxford: Macmillan Education.
- Wang, Y. (2010) Classification and SLA studies of passive voice. *Journal of Language Teaching and Research* 1 (6), 945-949.
- Wilson, C. B. (2006) Passing scores and passive voice: A case for quantitative analysis in writing assessment. *Exchanges: The Online Journal of Teaching and Learning in the CSU*.
- Xiao, R. (2007) What can SLA learn from contrastive corpus linguistics? The case of passive constructions in Chinese learner English. *Indonesian Journal of English Language Teaching* 3 (1), 1-19.

Appendix

Table A.1 Passive ratio of the 20 verbs in ICLE subcorpora (in descending order)

ICLE-EFL							
ICLE-FR		ICLE-GE		ICLE-KR		ICLE-SE	
CONCERN	81.75	INTEREST	69.44	INTEREST	64.29	SUPPOSE	86.84
SUPPOSE	78.13	SUPPOSE	64.10	CONCERN	40.54	DESCRIBE	59.57
INTEREST	71.74	CONCERN	49.18	SATISFY	33.33	INTEREST	54.84
CONSIDER	39.11	SATISFY	40.00	CONSIDER	28.31	CONSIDER	46.85
DESCRIBE	35.90	INVOLVE	29.73	INVOLVE	27.03	CONCERN	41.67
WRITE	21.82	CONSIDER	25.00	WRITE	18.06	SATISFY	39.29
SATISFY	21.43	USE	21.82	SUPPOSE	15.79	INVOLVE	37.93
INVOLVE	19.30	PROVIDE	13.95	USE	14.72	WRITE	20.00
USE	18.68	SEE	11.15	DESCRIBE	14.29	GIVE	13.15
FIND	17.03	DESCRIBE	11.11	NEED	9.12	SEE	9.75
SEE	14.63	GIVE	9.59	FIND	8.97	USE	9.62
NEED	13.48	SAY	7.74	GIVE	8.91	SAY	9.22
SAY	12.68	FIND	7.26	SEE	6.88	FIND	8.91
MAKE	11.25	NEED	6.70	SHOW	6.45	TAKE	7.96
PROVIDE	10.34	KNOW	6.67	MAKE	4.75	SHOW	7.20
GIVE	10.26	MAKE	6.40	SAY	4.67	MAKE	6.32
TAKE	9.28	SHOW	5.52	PROVIDE	4.60	PROVIDE	4.76
KNOW	3.90	TAKE	4.84	KNOW	4.09	KNOW	3.59
SUGGEST	3.85	WRITE	2.78	TAKE	3.39	NEED	2.33
SHOW	2.58	SUGGEST	0.00	SUGGEST	0.00	SUGGEST	0.00

