A corpus-based study of the passive voice in texts produced by Norwegian learners and native speakers of English

Eman B.A Sulaiman

ENG4191 – Master's Thesis in English Language
60 ECTS credits

The Department of Literature, Area Studies an European Languages

The Faculty of Humanities
University of Oslo

November 2022

2022

A corpus-based study of the passive voice in texts produced by Norwegian learners and native speakers of English

Eman B.A Sulaiman

http://www.duo.uio.no/

Trykk: Reprosentralen, University of Oslo

Abstract

Taking a corpus-based approach, the present study examines the passive voice usage by native speakers and Norwegian learners of English of different ages and educational levels, drawing on material from the LOCNESS, ICLE- NO, and TRAWL corpora. To accommodate the scope and time constraints, the focus is only placed on two types of passive: the central BE-passive and the central GET-passive.

The study sets out to uncover similarities and differences in the use of the BE and GET-passives in the (academic) writing by three students groups. Thus, qualitative and quantitative methods are adopted to investigate the frequency of the BE and GET-passive in each corpus and explore the factors influencing students to employ them in their writing. A further aim was to determine if native speakers and Norwegian learners of English make similar lexical choices, i.e. do native and non-native speakers use the same main verbs in combination with the passive auxiliaries?

The frequency analysis shows that Norwegian learners in ICLE-NO and the TRAWL corpora underuse the passive voice compared to native student writers in LOCNESS. Moreover, the frequency of the most frequent verbs associated with the passive voice confirms that the NS has more lexical variation than non-native Norwegian English learners. However, the three groups share some similarities in the use of the passive voice. For example, the three groups consider the BE-passive as the unmarked variant, while the GET-passive is the marked one. The same is true of the short and long passives; the short passive occurs significantly more frequently than the long passive in the NS and NNS writings.

Acknowledgments

I owe my gratitude to everyone who made this work possible. Without their contributions, I would not have been able to accomplish this thesis. Firstly, I wish to offer my sincere thanks to my supervisor, professor Signe Oksefjell Ebeling for her invaluable knowledge and expertise, constructive and encouraging feedback, support, proofreading, and guidance during my studies. I would like to take this opportunity to thank professor Hildegunn Dirdal for giving me access to the TRAWL corpus and making the corpus data available to me. She helped me come up with the thesis topic and guided me in my thesis proposal.

Finally, but definitely not least, my deepest love and gratitude go to my family for being encouraging and supportive with my ups and downs. I would especially like to thank my husband, Mohammed, for his patience, understanding, support throughout my academic life and for taking responsibility for our family. I would also thank my kids, Yousef and Mariam, for being understanding and for their constant encouragement and love.

Table of Contents

List of tables

List of figures

List of abbreviations

CL Corpus linguistics

CIA Contrastive Interlanguage Analysis

EFL English as a Second Language

ESL English as a Foreign Language

ICLE-NO The Norwegian part of the International Corpus of Learner English

L1 First language

L2 Second language

LCR Learner Corpus Research

LOCNESS Louvain Corpus of Native English Essays

NNS Non-Native Speakers

NP Noun Phrase

NS Native Speakers

TRAWL corpus Tracking Written Learner Language

1 INTRODUCTION

The present study is a corpus-based investigation of the use of BE and GET-passives by intermediate and advanced Norwegian-speaking learners of English (L2) in comparison with native speakers of English. The aim is to evaluate its use in terms of frequency, lexical association/variation, the absence and presence of the agentive phrase, and the factors that influence the use of one type over the other.

1.1 Purpose of the study

Passive voice has been defined as a complex and marked construction (Huddleston and Pullum, 2002; Carter and McCarthy, 2006) that poses difficulty not only for non-native speakers (NNS) of English but to native speakers (NS), as well. The use of the passive voice has always been controversial because of the negative attitude associated with it. It is claimed, "if you use passives: your writing will become weak, dull, vague, cowardly, bureaucratic, and dishonest" (Pullum 2014, 1), as it hides the agent doing the action. As a result, the passive voice has often been underused and replaced by its active counterpart (Folse 2009, Pullum 2014) by NSs and NNSs alike. The present corpus study investigates how frequently the passive voice is used in essays written by native speakers of English and Norwegian learners of English at different levels of education. The purpose is to find out if Norwegian learners utilize the passive voice in the same way as native speakers of English do or if they tend to underuse it in their academic writing.

The study is corpus-based, drawing on two corpora containing texts written by Norwegian learners of English at a university level (ICLE-NO) and by Norwegian pupils in lower secondary school (TRAWL); and one written corpus by native speakers of English at a university level and A-level (LOCNESS). The frequency and use of the passive voice is compared across the three corpora to reflect upon the differences in the passive voice frequency among learners in light of their level of education and age. Moreover, the comparison will also focus on lexical variation in the passive voice, i.e. to what extent NSs and NNSs select the same main verbs.

This thesis investigates two types of passives: the central BE-passive and the central GET-passive. Though BE and GET- passives have a similar syntactic structure, they have, among other things, distinct uses and connotations; while the BE-passive is prevalent in formal writing, the GET-passive is dominant in the spoken register. These distinctions and the distribution of each type in the corpora are provided in this study. I will also focus on different factors that might affect or restrict the production of the BE and GET-passive and the short and long passive constructions by NS and NNS. The result emerging from the study will aid both students and teachers to better understand the stylistic features of the passive voice in the academic genre and give insight into how NS and NNS writing styles vary.

1.2 Research questions

As already stated, this study mainly aims at identifying the similarities and differences of the use of the passive voice between native speakers of English and Norwegian learners of English, in particular, the relative frequency, the linguistics factors that may influence their use and lexical variation.

Thus, throughout this study, three research questions are pursed and investigated:

1. Do Norwegian learners of English use the passive voice to the same extent as native speakers do?

    1a. How frequent is the use of the central BE-passive vs. the central GET-passive?

    1b. How frequent is the use of the long vs. the short passive?
2. What factors influence the students to choose between BE and GET-passives, on the one hand, and between short and long passives, on the other?

3. Do Norwegian learners of English and native speakers have similar lexical variety when producing the passive voice?

An additional, overarching question, is to what extent age and level of education seems to play role in the Norwegian learners use of the passive voice. It will also be interesting to see if the result of the study may point to reasons for the students' choices.

1.3 Structure of thesis

The study is organized as follows:

Chapter 2 (Literature review) starts by describing the distinction between active and passive sentences. Then, it presents the passive voice in English and Norwegian, focusing on the two types of passive investigated in this study. Finally, the chapter ends by giving an overview of earlier relevant corpus studies of the passive voice.

Chapter3 (material and methods) presents the material used to elicit data, i.e. it describes the corpora from which the data are taken. It also explores the methodology applied, including data collection procedures, classification systems, and a preliminary frequency analysis.

Chapter 4 (analysis and findings) aims to provide answers to the research questions by discussing the study's findings. It firstly counts the frequency of the passive voice in general in each corpus and then examines the proportion of the BE and GET-passive and the short and long passives. It also explores the factors that affect the choice of each type. Eventually, the frequency of the most frequent verbs associated with the BE and GET-passive are provided and discussed.

Chapter 5 (conclusion) concludes and summarizes the most important findings and discusses some limitations of the study and suggestions for future research.

2 LITERATURE REVIEW

2.1 Introduction

The literature review conducted for the study is intended to provide a theoretical background for the research. Therefore, it has been organized to review the passive voice in English and Norwegian and look at some relevant previous studies of the passive voice. Section 2.2 sheds light on the active-passive alternation and discusses the view on using the two types of voice. Furthermore, since this study is a contrastive one, a description of the passive voice in English and Norwegian are given in section 2.3 and 2.4, respectively. Lastly, section 2.5 explains the notion of frequency analysis and discusses some relevant previous corpus-based frequency studies of the passive voice. Finally, section 2.6 discuses framework adopted in this study.

2.2 What is the passive voice

Quirk et al. (1985) define voice as a grammatical category that makes it possible to view the action of a sentence in two ways without change in the facts reported in the sentence. There are two basic voices to describe the relationship "between the participants and the event indicated in the verb" (Nida 1964, 200): active and passive voice. According to Shibatani (1988, 3), voice is "a mechanism that selects a grammatically prominent syntactic constituent -subject- from the underlying semantic function ('case' or 'thematic roles') of a clause." Celce-Murcia and Larsen-Freeman (2015) state that voice "uses a form of verb to tell us whether the subject is the actor or is acted upon." Thus, when the Subject is the agent executor of the action, the sentence has an active voice, while it has a passive voice when the Subject is the affected entity, and the agent may or may not be specified (Baker 1992, 102).

Moreover, Thornbury (2006,156) defines voice as "the way the relationship between the Subject and Object of the verb can be changed without changing the meaning of the sentence." That is, the meaning of the passive is grammatical rather than lexical, and it exists "to put the patient, i.e., the receiver or undergoer of an action, in subject position" (Celce-Murcia & Larsen-Freeman 1999, 347). Myers (1966), furthermore, compares active and

passive voice in a straightforward way, stating that constructions, like *I call*, are said to be 'active,' and constructions such as *I am called* are said to be 'passive.'

Broadly, the main distinction between active voice and passive voice is the role of the Subject referent in clauses that express an action. In the active voice, the Subject performs or causes the action expressed by the verb (Nordquist, 2016), and it is used to make a direct statement about the action, as William Strunk (1918) states, "use the active voice... The active voice is usually more direct and vigorous than the passive…". On the other hand, the passive voice is employed to make an indirect statement about an action where the focus is on the person or thing that experiences the action rather than on who performs it. While the active voice is the most common form in spoken interactions, the passive voice frequently occurs in written form, especially in scientific and academic texts (CelceMurcia & Larsen Freeman 1999, 354).

Although Huddleston (1976) agrees with this notional definition of the active and passive voice, he affirms that the difference in the semantic role of the Subject does not serve to define active and passive sentences. He points out that sentences such as *Everyone knows the answer* and *The answer was known by everyone* do not express the action of an actor on a patient. They are classified as active and passive, not because of their semantic roles, but because they show the same syntactic contrast as in the pair *John attacked Peter* and *Peter was attacked by John*.

2.3 Passive voice in English

Svartvik claims that "there is no agreement among grammarians as to what constitutes an English passive" (Svartvik 1966, 3). Although a large number of passive sentences consist of a form of the verb BE followed by the past participle, some passive clauses can also contain other copular verbs than BE, e.g., GET, BECOME, SEEM. According to Huddleston and Pullum (2002,1430), some passive constructions contain neither of these verbs, the so-called bare passives.

Nevertheless, the three most basic passive constructions in English are BE-passives (BE+ past participle), HAVE-passives (HAVE+ do+ past participle), and GET-passives (GET+ past participle), these are illustrated in the table 2.1. In the present thesis, only BE- and GET-passives are included in the analysis, as they are the most common passive forms and

predominate over the other forms. Therefore, I will be focusing only on the construction of these two forms in this section.

Table 2.1: Active and passive sentences

ACTIVE VOICE	PASSIVE VOICE
MARY WASHED THE CLOTHES.	BE-PASSIVE The clothes **were washed** by Mary. GET-PASSIVE The clothes **got washed** by Mary. HAVE- PASSIVE Mary **has the clothes washed.**

2.3.1 BE-passive

The basic passive voice structure in English (writing) is known as the reversed structure of its active counterpart (SVO), where the Object replaces the Subject position and the Subject is placed either in a *by*-phrase or omitted. The main Verb takes the past participle form by attaching its stem to the passive morpheme in the form of *ed* or -(*e*)*n* suffix (Ouhalla 1999,170). The passive construction is typically preceded by a free morpheme in the form of the passive voice auxiliary BE. This auxiliary can be found in different forms: *is, are, am* (denoting present tense), *was, were* (denoting past tense), or *be*, *being*, *been* (denoting a lack of tense in passive non-finite clauses).

Radford (1997) adds that passive sentences may contain a *by*-phrase in which the complement of *by* thematically corresponds to the expression functioning as the Subject in the active counterpart of the sentence. Moreover, the passive expression, which serves as the Object of an active verb, surfaces as the Subject in the corresponding passive sentence. The following are examples of an active sentence (2.1a) and its passive counterparts (2.1b and 2.1.c):

(2.1) a. Bill washed the dishes .

b. The dishes **were washed** by Bill. (With agent)

c. The dishes **were washed**. (Agentless)

Sentence (2.1a) is an active sentence with a structure of Subject *Bill* (the agent/doer) + verb in active voice *washed* + the direct Object *the dishes*. The Subject in this sentence is the agent doing some action, hence the sentence's focus. Sentence (2.1b) is a passive sentence in which the Subject *the dishes* is the receiver of the action, or patient, followed by the auxiliary verb+ past participle, *were washed* + *by*+ the agent *Bill*. The passive voice has reversed the active voice in that the Object of the action becomes the subject and, thus, the primary focus of the sentence.

As can be seen in (2.1b), the agent appears as a complement of the preposition *by*, which is the most common preposition to demonstrate the agented passive. The verb phrase consists of the auxiliary BE carrying the tense inflection, followed by the main verb *washed* in the past participle. Although the active and passive constructions in examples (2.1a) and (2.1b) are syntactically different, they still present the same information in the sense that if (2.1a) is true, then so is (2.1b), and conversely. The problem is to know in which situations the passive voice is preferred. The most common uses will be discussed later in this chapter and in chapter (4).

In (2.1c), the focus is on *the dishes*, and the agent is ignored and omitted. In this case, example (2.1c) does not have an exact active counterpart since it says nothing about who was responsible for washing the dishes. Huddleston and Pullum et al. (2002, 1428) also take the view that agentless passives have no exact active counterpart.

2.3.2 GET-passive

The GET-passive has almost the same syntactic structure as the BE-passive; i.e., GET is followed by a past participle and an optional *by*-phrase. In line with the BE-passive, "the GET-passive construction presents a process or event as undergone by the subject" (CelceMurcia and Larsen-Freeman 1999, 354). Hence, the resemblance between the BE-passive and GET-passive makes them interchangeable in many contexts. Still, they are not interchangeable in some circumstances, such as when the main verb refers to a state one, see section (4.3.1.1), (Quirk et al. 1985, 161, Carter and McCarthy 1999, 51-52, and Huddleston and Pullum 2002, 1442). Similarly, Alexiadou (2005, 17) argues that "the get-passive is not permitted with stative verbs and verbs that do not allow for the subject of the construction to be interpreted as affected." That is to say, the GET-passive is restricted to dynamic verbs that donate an action such as *paid, fired, killed, cut, or* arrested, as in the examples (2.2-2.3) below from Huddleston and Pullum et al. (2002,1442):

(2.2) It **was/*got believed** that the letter was a forgery.

(2.3) He **got arrested**.

Furthermore, the semantic and pragmatic implications differ between BE- and GETpassives. Firstly, the GET-passive implies a (sudden) change, while the BE-passive indicates a result. Huddleston and Pullum (2002,1442) note that "*get* tends to be preferred over *be* when the subject-referent is seen as having an agentive role in the situation or at least having some responsibility for it." Therefore, *Tom got killed* depicts a consequence of, for example, fighting in a battle, where *Tom* plays a more active role. While the version with BE, *Tom was killed*, indicates that someone killed *Tom* without *Tom* having triggered this. Consequently, GET-passives have been said to be agentive or have subject responsibility (see 4.3.1.3).

Additionally, the use of the GET-passive is more common with verbs implying negative effects or those referring to unfortunate events, such as *getting killed, arrested,* and *hurt* (see also 4.3.1.2). Carter and McCarthy (1999) found in their spoken corpus that almost 90% of GET-passives refer to adversative meaning while fewer than 5% have beneficial meanings. They state that GET-passives are a "state of affairs that is signaled contextually by the conversational participants as unfortunate, undesirable, or at least problematic" (ibid, 49). Additionally, GET-passives seem to be primarily limited to informal registers, so we find them more frequently in spoken language than in writing. For example, Biber et

al. (1999, 476) point out that in the Longman corpus, the GETpassive "occurs only in conversation, except for an occasional example in colloquial fiction." Nevertheless, the GET-passive construction has increased dramatically in written English over the last few decades, according to Leech et al. (2009, 156).

Furthermore, Huddleston and Pullum (2005, 245) state that GET, in contrast, to BE, is not counted among the primary auxiliary verbs since it does not display the "NICE" properties (Negative, Interrogative, Code, Emphasis). Thus, GET cannot precede *not* in negative sentences (2.4) or precede the subject in an interrogative sentence (2.5). Likewise, in example (2.6), the GET-passive cannot be used to substitute the main verb (the "code" test) or to carry the emphasis, as in (2.7). Hence, GET-passives must be preceded by the dummy operator (do) *when they occur* in these cases.

(2.4) She was not arrested.
*She got not arrested.
She did not get arrested.

(2.5) Was she arrested?
*Got she arrested?
Does she get arrested?

(2.6) Was she arrested? She was.
*Got she arrested? She got.
Did she get arrested? She did.

(2.7) She was arrested.
*She got arrested.
She did get arrested.

2.3.3 Short and long passive

An important criterion often used for categorizing passives is the presence or absence of a *by*-phrase to specify the agent of the action. When the *by*-phrase occurs in the sentence, the passive is called the long passive, as in example (2.1b) above, whereas the short passive is when the Subject is absent and the verb is not followed by a *by*-phrase, e.g. (2.1c). In principle, the long passive, as mentioned earlier, can be replaced by an active clause conveying the same meaning. However, in some situations, the long passive voice seems more appropriate (Biber, Conrad, & Leech, 2003). For instance, long passive constructions

are used when the agent is an indefinite noun phrase conveying new information; see below in this section and see also (4.3.2) for further discussion.

Often, long and short passives are called agentful and agentless passives, respectively (Biber et al. 1999, 935). However, based on Coetzee (1980), the short passive should not be treated as a short version of the long passive but as agentless sentences. Studies like Granger (1983) and Givón (2001) have observed that short or agentless passives are the most typical type of passive in English. According to Svartvik's (1966) analysis of corpus data, four out of every five English passive sentences are short passives. Moreover, short passives are more frequent in scientific writing, as observed in Coetzee's study (1980), because they allow the readers and the author to hide and rediscover the hidden agent and "enable us to LEAVE OUT something that would be obligatory in the active, namely a main clause subject." (Coetzee 1980).

Yet, there are several situations where long passives are still required. According to Yannuar, et al. (2014, 1401-1402), there are some principles for choosing long passives in writing:

1. The information-flow: where the new information is preferred to be at the end of a clause.
2. End-weight: in which the agent does not hold up the processing of the rest of the clause.
3. The long passive places initial emphasis on an element of the clause, which is the topic or theme. That is, it allows the Object agent to have less attention from the readers and somehow lack of responsibility.

In addition, Raimes (1998) mentions in her book "*How English Works*" some other common uses of long passives**.**

1. When the agent is not a person but an inanimate object: *the alarm is triggered by photomagnetic cells.* (Raimes 1998, 127-29).

2. To put focus on the receiver of the action while also giving credit to the agent: *St.Paul's Cathedral was designed by Christopher Wren* (ibid).

3. When the structure of the sentence or the relation between two sentences determines that the new information should come last: *the vice precedent wrote a report. That report is being studied by all the company officers* (ibid, 128).

Moreover, according to Shintani (1979), the writer tends to express the agent in some instances, such as when: (these instances will be discussed and illustrated with examples in chapter 4).

1. The agent is expressed when it is a proper name indicating an artist, an inventor, a discoverer, or an innovator.

2. The agent is expressed when it is an indefinite noun phrase conveying new information that the speaker/writer thinks is important enough to mention.

3. The agent is an unexpected inanimate noun.

2.3.4 Passive transitivity

In order to passivize an active clause, the verb should be transitive, provided that a direct object can follow transitive verbs. Thompson et al. (2013, 1) demonstrate that "a transitive event is one involving two participants: an agent, the 'doer' of the action, and a patient, the person or thing that 'undergoes' the action. In English, such events can be described in the active voice or passive voice." Pinker (1989, 136) argues that transitive verbs may undergo passivation because they include both an agent and theme/patient argument. However, according to Chomsky (as cited in Crawford 2012, 18), not all transitive verbs can be passivized; for example, stative verbs like *cost*, *weigh, resemble*, and possessive *have* may not undergo passivation (2.9-2.11), as the passive is meant to show that the Subject of the passive construction is somehow 'affected' by the verb, and it is not the case when the verb is stative. Intransitive verbs, by definition, are never followed by a direct object nor undergo passivation. The active sentences and their ungrammatical passive counterparts are illustrated below (2.9-2.11), (ibid).

(2.9) The book costs 20$.
* 20$ **is costed** by the book.

(2.10) I had flu.
* Flu **was had** by me.

(2.11) John resembles his father.
*His father **is resembled** by John.

2.3.5 Passive gradient

As already mentioned, although the passive voice has no clear-cut definition, it is broadly defined as BE or GET followed by the main verb in past participial form and sometimes also a prepositional phrase beginning with *by* (Quirk et al. 1985, 159f). However, the ambiguity of the BE+ past participle construction, in particular, has made the boundaries between passive and non-passive constructions fuzzy since this broad definition can be extended to include numerous constructions that are not considered passive (Quirk et al. 1985, 167). Thus, the term *the passive gradient* has steadily been used among grammarians and linguistis to refer to the sliding scale that ranges from constructions that clearly function as passives to those which may seem to be passive on the surface but are, in fact, not regarded as such (ibid, 167- 171).

Quirk et al. (1985, 167) propose a scale to distinguish between different categories of passive; this scale is divided into central passive, semi-passive, and pseudo-passive sentences[1], see table (2.2).

1. Quirk et al.'s terminology corresponds to the terminology used in the present study.

Table 2.2: The passive gradient by Quirk et al. (1985,167)

Categories	Central features	Examples
Central passives	They have active counterparts and only verbal properties. The agent *by*-phrase can be both expressed and left out.	The book **was written** by Sarah.
Semi- passives (adjectival passive)	They have active counterparts and both verbal and adjectival properties.	Leonard **was interested** in linguistics.
Pseudo-passives	they are only passive in function BE+ *ed* participle, but have active meaning and they do not have active counterparts nor agent *by*-phrase.	The building is already **demolished**.

Central passive refers to passive constructions that have clear relationships with an active correspondent; that is, they have active counterparts in which the agentive *by*-phrase can be both expressed and left out and have only verbal properties, for example:

(2.12) This violin **was made** by my father.

(2.13) My father made this violin.

(Quirk et al. 1985, 167)

Sentences that are described as semi-passive contain participles that have both verbal and adjectival properties, and the reason for considering them as verbs is because they have active counterparts. Semi-passive sentences can be modified by *very*, *quite*, *more*, *rather*, *etc*., and allow for the auxiliary BE to be replaced by a copular verb such as *seem* or *feel*. Besides, they can coordinate the participle with an adjective, such as *keen* (ibid).

(2.14) Leonard **was interested** in linguistics.

(2.15) Leonard **seemed very interested** in and **keen** on linguistics.

(Quirk et al. 1985, 168)

Finally, Pseudo-passive sentences are considered passives solely because they have the same structure as passive sentences, which include verb+ *-ed* participles. Unlike the central and semi-passive, this group does not have an active counterpart nor an agentive by-phrase (Quirk et al. 1985, 169). Their strong adjectival characteristics allow the verb BE to be replaced by other copular verbs such as *seem*, *become*, *remain*, *etc*.

(2.16) In 1972, the Democrats **were defeated.**

(Quirk et al. 1985, 170)

Example (2.16) above can be read in two ways, via a *dynamic* reading or a *statal* reading (Quirk et al. 1985, 170). In the former reading, the clause is considered a passive in which *Someone defeated the Democrats*, if there is a possibility to derive an agentive *by*-phrase from the context, the sentence is then considered central passive. In the latter interpretation, the verb *were* is considered a copular verb denoting a state resulting from the defeat rather than to the act of defeating itself, that is, *The Democrats were in a state of having been defeated*.

Next, in her corpus-based investigation of the passive in the spoken register, Granger (1983, 81-190) also breaks down BE+ participle (BE *Ved*) constructions into several categories; namely, passives, adjectival pseudo-passives, verbal pseudo-passives, mixed be+ Ved combinations, usually- passive category, peripheral combinations, and stative combinations. In comparison with Quirk et al. (1985), she preserves the term passive for central passives, "which stand in direct alternation to a semantically equivalent active verbal group" (Granger 1983, 108). In addition, she distinguishes between two categories of pseudo-passives: adjectival and verbal. Peripheral and mixed combinations, on the other hand, "share all characteristics of passives but whose active counterparts are far less common." Granger (1983, 105-115) further points out that the first three constructions are considered passives, whereas the latter four categories are borderline cases, and sometimes they can be put into the first category.

Examples (2.17)- (2.24) illustrate the seven categories, respectively, all from Granger (1983, 108ff).

(2.17) That attitude **was maintained by the government** in the further nine days of debates in the Lords. (Passive)

(2.18) We knew quite well that the tsarina **was devoted** to Russia and to her Russian subjects. (Adjectival pseudo-passives)

(2.19) She's been rather elusive as far as I'**m concerned**, so I don't really know her. (Verbal pseudo-passives)

(2.20) I**'m** very **interested** in poetry. (Mixed BE+ Ved combinations)

(2.21) I feel we**'re** all **faced** with this problem. (Usually-passive category)

(2.23) I**'m** fairly closely **connected** with that work. (Peripheral combinations)

(2.24) But I have these two houses that **are built** on to the next door's back garden sort of thing. (Stative combinations)

The fuzziness and overlap between passive and non-passive categories is represented by Granger (1983, 107) in the form of three interlocking circles, in figure (2.1). The non-overlapping parts illustrate three clear-cut categories: passive, adjectival pseudopassive, and verbal pseudo-passive, whereas the last four categories, i.e. statal, usually passive, peripheral, and mixed, are borderline categories placed in the overlapping area (Granger 1983, 107).

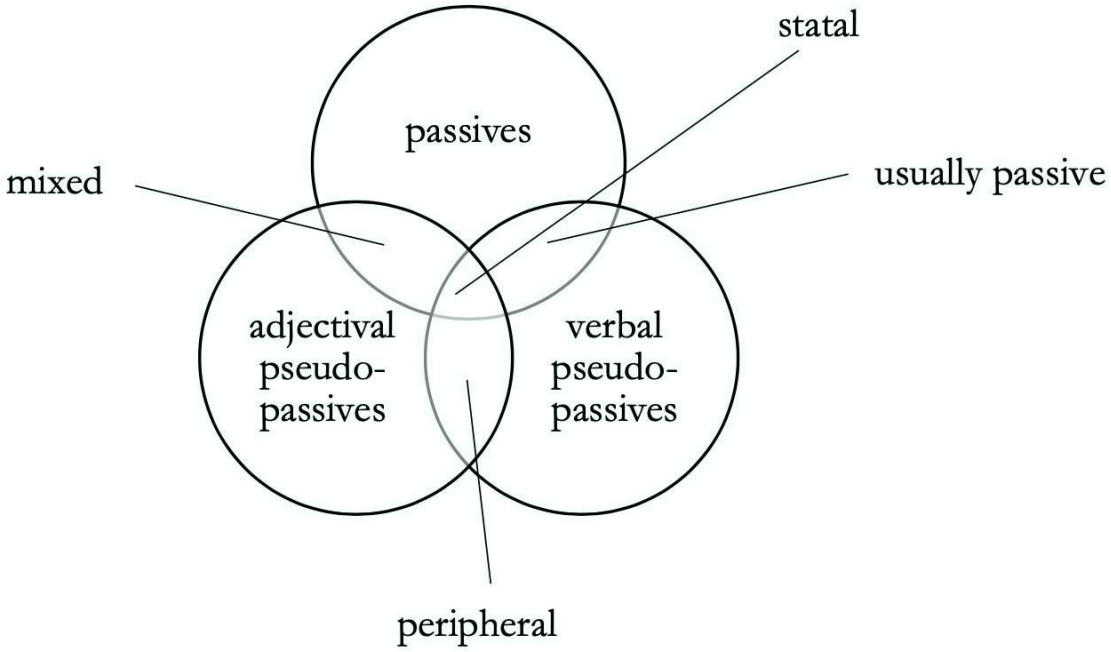


Figure 2.1: Granger's representation of interaction between passive and non-passive categories (Granger 1983, 107).

Later, Huddleston et al. (2002) and Pullum (2014) identified other types of passive constructions besides the basic passive construction (BE+ past participle) and adjectival passives, (which are called semi-and pseudo passives by Quirk et al. 1985. These passive constructions are as follows:

(1) prepositional passive: where the Subject in the passive structure corresponds to the Object of a preposition in the related active structure. (*He was laughed at by his friend- His friends laughed at him*).

(2) Bare-passive clauses: this category refers to non-finite clauses that contain only a Subject and the past participle of the verb. Thus, the auxiliary BE or GET is omitted. (*My house wrecked by a tornado is something I don't ever want to see*).

(3) Embedded passives refer to the passive structures embedded in active clauses. (*The government had the case investigated by the police*).

(4) Concealed passives refer to passives with gerund instead of the past participle as the head. (*This house needs painting*).

2.4 Passive voice in Norwegian

Though most of the world's languages have at least one construction called passive, the differences between the various constructions are quite significant (Siewierska 1984; Keenan 1986). This study is a contrastive interlanguage study where the data are taken from L1 English and L2 English by L1 Norwegian students. Thus, to give a clear picture of the passive voice usage between the two groups, the similarities and differences regarding passive voice constructions in English and Norwegian are discussed in this section.

In English, the passive is formed only periphrastically by the auxiliaries BE, GET, or one of the more marginalcopular verbs such as BECOME and SEEM, followed by the past participle form of the verb; see section 2.3 for further information. In Norwegian, on the other hand, there are two basic ways of forming the passive voice: a periphrastic one and a morphological one (table 2.3 below). The periphrastic passive is constructed by one of these three auxiliaries, BLI, VÆRE, and FÅ, followed by the past participle of the main verb.

The auxiliaries, the BLI (BECOME) and the VÆRE (BE), are the main auxiliary verbs used to construct the periphrastic passives. Although the difference between the VÆRE-passive and BLI-passive is semantic in nature, the Norwegian VÆRE-passive typically indicates a lasting activity or state when it precedes non-mutative verbs while focusing on the end result when the past participles are dynamic. The BLI-passive, however, is used for describing single events with emphasis on the event itself (Faarlund et al.1997, 525).

The semantic differences are exemplified below (these examples are taken from Faarlund et al. 1997, 525):

(2.25) a.Veien **er åpnet**. = 'The road is in a state of being open.'
(*Lit:* 'The road is opened.')

b. Veien **blir åpnet**. = 'The road is being opened.'
(*Lit:* 'The road becomes opened.')

(2.26) a. Kjolen **var renset** = 'The dress was in a state of being clean.'
(*Lit:* 'The dress was cleaned.')

b. Kjolen **ble renset** = 'The dress was being cleaned.'
(*Lit:* 'The dress became cleaned.')

The last auxiliary to form the periphrastic passive is FÅ (GET). In line with the VÆRE-passive and the BLI-passive, the FÅ-passive is also followed by a past participle of the main verb. However, the FÅ-passive is a controversial passive type. It is considered a passive or passive-like construction in some studies, as in Ryen (1990), while in other studies, it is rarely treated as a part of the Norwegian passive system (Faarlund et al. 1997, 848), although Ryen (1999, 194) in the table below does so. Moreover, the FÅ-passive carries some restrictions to consider it passive rather than a modal construction. For example, the FÅ-passives are limited to transitive verbs and should be followed by an agentive *av*-phrase '*by*-phrase' (Faarlund et al. 1997, 848).

Table 2. 3: Inflection of the morphological and periphrastic passive in English andNorwegian (Ryen 1999, 194)

	Morphological	*Periphrastic*		
	s-passive	*bli-passive*	*være-passive*	*få-passive*
Present	arbeides	blir arbeidet	er arbeidet	får arbeidet
	work$_{\text{s-passive}}$	*becomes worked*	*is worked*	*gets worked*
Past	arbeidtes	ble arbeidet	var arbeidet	fikk arbeidet
	work$_{\text{s-passive}}$	*became worked*	*was worked*	*got worked*
Present perfective	*n/a*	er blitt arbeidet har blitt arbeidet	har vært arbeidet	har fått arbeidet
		has/have become worked	*has/have been worked*	*has/have got(ten) worked*
Past perfective	*n/a*	var blitt arbeidet hadde blitt arbeidet	hadde vært arbeidet	hadde fått arbeidet
		had become worked	*had been worked*	*had got(ten) worked*
Present future	skal arbeides	skal bli arbeidet	skal være arbeidet	skal få arbeidet
	shall work$_{s\text{-passive}}$	*shall be worked*	*shall be worked*	*shall get worked*
Past future	skulle arbeides	skulle bli arbeidet	skulle være arbeidet	skulle få arbeidet
	should work$_{\text{s-passive}}$	*should become worked*	*should be worked*	*should get worked*

The morphological form is the second way to construct the Norwegian passive, and it has no counterpart in English. It is simply formed by adding the suffix-*s* to the verbal stem, the so-called *s*-passive (Faarlund et al. 1997, 507-509). The distinction between the morphological and periphrastic passive voices is explained as a distinction between general and specific events (Thorell 1973, 135; Western 1921, 159–61). Faarlund et al. (1997) point out that there is a semantic difference between the *s*-passive and the periphrastic forms of the Norwegian passive. The *s*-passive usually tends to be used after modal auxiliaries, and it is applied to talk about something general in nature or ongoing events rather than focusing on a single or concrete event as the periphrastic passive does (ibid, 514-515). Additionally, many grammarians agree that the *s*-passive is the most common type in perceptions, norms, rules, and commands. (e.g. Hansen 1967, 147; Engdahl 2006, 24; Thorell 1973, 135). In line with the *s*-passive, the English Get-passive is more likely to be used in commands such as (*get your hair cut)*. Examples of the periphrastic and the morphological passive are presented in table 2.3 above.
In his study, Hovdhaugen (1977: 36ff) investigates the distribution of the morphological (*s*-passive) and periphrastic passive constructions. His investigation shows that the morphological form is dominant, whereas periphrastic constructions are less often used. Accordingly, the frequent use of the *s*-passive by Norwegian learners may affect the frequency of the English passive voice in general and the GET-passive in particular. Therefore, one would expect Norwegian students to substitute the Norwegian *s*-passive either with its active counterpart or the GET-passive when writing English texts.

Nonetheless, the morphological passive in Norwegian has some structural restrictions regarding the past tense, as it is used only with regular verb classes that end with -*e* in the past tense, such as *arbeide* in table 2.3. Irregular verbs cannot, with few exceptions, be used in the past tense with the morphological passive, as in (**skrev-s*). Moreover, the morphological passive is also absent in both the past and present perfective. In the periphrastic constructions, in contrast to the morphological passive, it is the auxiliary verb that is inflected for tense rather than the main verb; therefore, as table 2.3 displays, the tense paradigm in both languages is completed.

Moreover, while the typical English passive is mainly formed with a transitive verb (as mentioned in section 2.3), the Norwegian passive maybe formed with an intransitive verb.

Hence, both the morphological and periphrastic forms can be used in impersonal passive constructions. The agent of the sentence in the Norwegian passive, similar to the English passive, is optionally presented in the form of a prepositional phrase with the preposition *av* (by). However, the Norwegian impersonal passive presupposes human action, which means that it is restricted to include only verbs that express human action and are durative in nature (Hovdhaugen 1977, 24), as in example (2.27).

(2.27) Men arbeides det her?
But work-S it here?
'Do People work here?'

(Faarlund et al. 1997, 854)

As we can notice from example (2.27), the agentive prepositional phrase is not expressed overtly through an *av*-phrase because it is generally considered unnatural in these constructions (Hovdhaugen 1977, 24).

Thus, from a contrastive perspective, both similarities and differences between English and Norwegian passives have been outlined above. They are similar in that both languages have the periphrastic passive, while they differ in that the Norwegian passive also has a morphological variant. Another important distinction between the languages is related to impersonal passive constructions. Norwegian may form an impersonal passive with an expletive Subject and an intransitive verb in contrast to English, where only transitive verbs can form a passive.

2.5 Relevant previous studies of the passive voice

This section first mentions the benefits of using corpus linguistics as a method to investigate passive voice frequency. Then, it presents three influential corpus-based frequency studies, sheds light on their methodologies and analysis, and situates my own study in relation to them. The section closes by presenting the main assumptions regarding the passive voice and mentions the framework adopted in this study.

2.5.1 Frequency study of the passive voice

The passive is a vast topic and has always been an object of study for grammarians and linguists. Most of them have extensively studied the passive voice with respect to its frequency (where the focus has mainly been on its relationship to the active voice), the frequency of agentless and agentive passives, and also on examining how (frequently) NS and NNS of English use the different types of passive (BE-passives, GET-passives, HAVE-passives).

To investigate voice in terms of frequency, researchers have relied heavily on corpus linguistic methods as it consists of real language samples (naturally-occurring data). For example, Granger (2013, 6) notes that corpus linguistics provides "data and tools to give much more precise insights into passive use." Thus, instead of relying only on native speakers' intuition or elicited data which often contains little information on authentic language use, corpus linguistics enables us to access authentic language usage via a vast number of software programs and find out what language is really like.

Significant studies have employed frequency analysis to examine learner language in its own right. An advantage is to describe the passive as it is used "rather than in relation to target language norms" (Ellis & Barkhuizen 2005, 93). For example, concerning the frequency of GET-passives in written language, Collins (1996) points out that GETpassives are less common in formal language settings. Likewise, Mindt (2000) claims that the GET-passive is more dominant in the spoken medium. Similar findings are reported in the study by Leech & Svartvik (1994, 330); they note that GET-passives are only found in the informal register and constructions without an agent. Similarly, Huddleston and Pullum (2002, 1442) report that GET-passives are generally much rarer than BE-passives and tend to be avoided in the formal register. However, Schwarz (2015) investigated the decline in the use of the base BE-passive and the increase in the informal GET-passives in written language. She found that the GET-passive frequencies were the same in the last two years, whereas the BE-passives frequencies showed a 25% decline.

Moreover, in investigating passives collected from texts from various registers, Svartvik (1966) summarizes that the passive is relatively more frequent in written than spoken language and in informative rather than imaginative prose. Additionally, it has been argued that the short passive accounts for a large number of the occurrences of BE- and GET-

passives both in spoken and written language. For example, Quirk et al. (1985, 164) note, "approximately four out of five English passive sentences have no expressed agent." Jespersen (1933) and Palmer (1965b) also observed similar findings regarding the high frequency of short passives in English.

Many corpus-based studies of learner language have investigated the production of passive voice to compare the frequency of passive voice in learner language to native speaker language on the one hand and to find common L2 problems on the other hand. Many of them have reported the underuse of passive voice among learner populations. For example, in his study, Hinkle (2004) discovers consistent underuse across the six learner groups he analyses (Chinese, Japanese, Korean, Indonesian, Vietnamese, and Arabic). However, the degree of underuse varies significantly across the learner populations. Granger (2013) has also observed that many who has English as a second language (ESL) and as a foreign language (EFL) underuse the passives in their production. Gilquin (2008, 6) explained that learners' preference for the unmarked rather than the marked option could be one of the reasons for this underuse, and she goes further to say that the underuse of the passive among learner groups seems to be a universal feature of interlanguage.

In the following three subsections I will outline three previous studies that are particularly relevant for my own investigation.

2.5.2 Moreb (2016)

Moreb (2016) examined the frequency of BE-passive, GET-passive, and HAVE-passive in samples from freshman academic textbooks. The data was collected from four of the five common freshman academic textbooks from the General Education Programs [GEP], namely English composition, history, psychology, and biology. The study aims to evaluate whether the frequency of passive voice usage between the content areas is significantly different. To collect the data, Moreb counted all the examples of BE-passives, GET-passives, and HAVE-passives in each textbook sample and all the verb occurrences across the different academic textbooks. Later, the percentage of passive voice verbs and the total number of verb occurrences in each subject were compared (Moreb 2016, 38). Then, the researcher identified and compared the frequency of the basic passive voice tenses and aspects between the academic textbooks and the frequency of passive voice verbs by subject texts.

The results showed that the most frequent passive form was the BE-passive, with a frequency of 201 of the 205 passive voice forms. This is not surprising as the BE-passive is considered more formal and common in written language than the GET-passive and HAVE-passive. The highest number of BE-passives occurred in history and biology textbooks, with occurrences of 67 and 78, respectively (ibid, 66). Additionally, the most frequent passive voice tenses in the textbooks varied between the present and past tenses. While the simple present tense occurred more frequently in biology textbooks, the past tense dominated in history textbooks. In addition, a chi-square test showed that the frequencies of the four academic genres were statistically significantly different ($p < 0.05$) (ibid, 68).

5.2.3 Gilquin and Granger (2021)

The study by Gilquin and Granger (2021) has two primary objectives. Firstly, it intends to compare EFL and ESL learners to native speakers of English in terms of passive frequency, lexical preferences, and phraseological sequences, and secondly, to compare non-native varieties with each other. The researchers have adopted a lexico- grammatical perspective to study different varieties of L2 English produced by learners from eight L1 backgrounds. In four of them, English was considered an EFL language, whereas the other four represented ESL-like populations. This study seeks to determine whether employing passive constructions in native and non-native writing is purely grammatical or is attributed to the fact that some verbs are more attracted to the passive voice than others, the so-called phraseological effect (ibid, 72).

The data in this study comes from the latest version of the International Corpus of Learner English (ICLEv3; Granger et al. 2020), which contains argumentative texts written by university students from different L1 backgrounds. It was collected by extracting all the occurrences of lexical verbs tagged as past participles and preceded by the lemma BE, allowing for the possibility of one word between these two elements, as well as all cases where BE is followed by a verbal form tagged as an adjective (e.g., *be concerned*). The analysis shows underuse of the passive by the EFL and ESL learners, especially among Serbian, Norwegian, German, Hong Kong, and Korean learners, with a relative frequency of 221,2 per 100,000 words in the ICLE sample, as compared to 288,8 in LOCNESS

(Gilquin and Granger 2021, 85). This result shows that differences in the degree of exposure to the target language, a so-called usage-based view of language acquisition, do not necessarily lead to differences in the passive frequency. It is worth mentioning that my study differs from Gilquin and Granger study in that it focuses on Norwegian learners only and investigates only type of the passive, i.e. central passive voice, defined by Quirk et al. (1985). Thus, an example like the one mentioned above (*be concerned*) is not considered a central passive; hence it is removed from my data.

2.2.4 Gustafsson (2014)

Gustafsson runs a corpus-based study to examine the grammatical construction of BEpassives and GET-passives and their usage by Swedes writing in English in terms of frequency and semantics. In this study, the researcher focused on two passive patterns, namely, long passives and short passives, as well as adversative and non-adversative GET-passives. The data was extracted from four corpora, two of which were compiled for this study: SWENC (the Swedish English Newspaper Corpus) and ESC (Blogs in English by Swedes Corpus).

The result from these two corpora was put in a broader perspective by comparing them with the result from two native English corpora, namely, the press sub-corpora of the native English Frown 1991 (American English) and F-LOB 1991 (British English) corpora. This enabled the researcher to compare the BE-passives and GET-passives across crosslinguistic genres and between Swedish speakers of English and native speakers (Gustafsson 2014, 24).

After counting all the BE-passives and GET-passives, the result shows that the BEpassives occur considerably more frequently than the GET-passives in all four corpora. It indicates that Swedes writing in English use passive constructions to a similar extent as native English speakers do within the news genre. However, the GET-passives appear more commonly within the blog genre, given that this genre includes more informal language than the online newspaper (Gustafsson 2014, 14). Moreover, it was reported that the long BE-passives were dominant in Frown and F-LOB corpora, whereas they were less frequent in BESC and SWENC corpora. Regarding the GET-passive, most GET-passives in BESC and SWENC were adversative.

Most previous studies in the field based on a comparison of the passive voice production between L1 and L2 learners or between several non-native varieties (EFL and ESL) say something about the underuse, overuse, misuse, the proportion of the passive voice and the influence of the L1 on the L2. However, as far as I know, few of these have focused on comparing the production of passive voice between two groups of the same L2 but at different levels of education and of different ages. Therefore, the present study will not only compare L2 production to native speakers of English but also aims to contribute to this field by investigating the writing of L2 Norwegian university students in ICLE-NO and L2 Norwegian school pupils in the TRAWL with respect to the frequency of the passive and lexical variation in the passive voice. This will help track the developmental features in acquiring and using the passive voice by the three groups.

Based on the findings from previous studies, a few general hypothesis concerning the frequency of the passive voice in the L2 English production by Norwegians can be stated:

1. The frequency of the passive voice in the writing of Norwegian learners will be lower than in the corresponding writings of native speakers.

2. The BE-passives will be significantly more frequent than the GET-passives in Norwegian writings.
3. The frequency of the occurrence of agentless passives will be higher than the agentful passive.

2.6 Framework adopted

The chapter has pointed out some essential theoretical background regarding the passive voice and has gavin an overview of the construction of passive voice in English and Norwegian. Finally, the chapter has reviewed some of the relevant corpus studies on the passive voice.

Regarding the framework adopted for the analysis of the passive voice in the present thesis, I will mainly follow Quirk et al. (1985) in that, among other things, the main verb has purely verbal properties, has an active counterparts and the auxiliary can be replaced by a lexical

copular. A more detailed account about how the central passive was identified in the data is given in section 3.4.1of the method chapter.

3 MATERIAL and METHODS

3.1 Introduction

In the first part of this chapter I present the methodology used to collect the data. The second part presents the three corpora used and argue for their comparability. Finally, the last part explores the software employed in retrieving data from the corpora and go through the process of extracting the data and the main criteria for classifying the passive voice as central passive.

3.2 Method

In this study, both quantitative and qualitative methods are adopted to collect and process the frequency and use of the passive voice. A large proportion of the study is based on frequency information provided by the material in the corpora. The counting of the occurrences of BE and GET-passive constructions and analyzing their frequencies is, thus, a quantitative approach. The qualitative part takes a closer look at the type of the passive in terms of the presence or absence of *by*-phrase, their lexical association with main verb, and the factors influencing the choice of the passive type.

The following sections describe in detail the methodological issues considered in the present study. Sections 3.2.1 and 3.2.2 define the notion of corpus linguistics and learner corpora as a methodology, respectively, and discuss some of their benefits in the study of language.

3.2.1 Corpus linguistics (CL)

A corpus refers basically to a systematic computerized collection of authentic texts from both written and spoken language to reflect the real usage by native or non-native users of a particular language. The term corpus has been offered various definitions, some of which will be mentioned in this section. For example, Gries (2017) defines a corpus as "a machinereadable collection of (spoken or written) texts that were produced in natural communication settings." That is, all the texts in the corpus are produced in real situations and natural settings without any restrictions on length or any massive edition (7-9). According to Johnson & Johnson (1999), a corpus is "a large computer-held collection of

texts (spoken, written, or both) collected together to stand as a representative sample of a language or some part of it (89-90). They further state that "Corpora provide easily accessible and accurate data, useful to descriptive and theoretical linguistics." (1999, 8990). As a result, in the last few decades, corpus linguistics has developed to become a field of interest with an immense number of studies on the English language by many researchers. To begin with, corpus linguistics started its use within the English language, but later it expanded to include many other languages and fields.

Although the use of corpora remains a source of debate to this day, their contribution to linguistics in general, and English linguistics in particular, is now widely acknowledged. Corpora have proven their role in facilitating the analysis of a large scale of real examples of language in use. It also guarantees more reliable and accurate data since they give us an actual picture of what language is really like rather than relying on our intuition as a main source of data. Hunston (2002) notes, "A corpus essentially tells us what language is like, and the main argument in favor of using a corpus is that it is a more reliable guide to language use than native speaker intuition is" (20).

The term corpus linguistics (CL) has been controversial and has been used to refer a tool, a method, a theory, a methodological approach to studying a language, a discipline, a theoretical approach, a paradigm (theoretical or methodological), or a combination of these. For the purpose of this study, the definition given by Bowker and Pearson (2002) will be adopted "an approach or a methodology for studying language use" (9). Similarly, Gries (2006a) views corpus linguistics as a paradigm stating that "over the past few decades, corpus linguistics has become a major methodological paradigm in applied and theoretical linguistics" (191).

The advantages of using corpora in linguistic research are many and well-known. Johnson and Johnson (1999) stress the importance of corpus linguistics in collecting quantitative data, such as frequency counts and statical measures, noting that "A corpus… may also be used to calculate the frequency of occurrence of items and, as repositories of actual instances of language use, they have a place in language teaching textbook design" (9890). Meyer (2002) claims that corpus is a helpful method in exploring language and developing linguistic studies:

> [The use of] corpora [has] numerous uses, ranging from the theoretical to the practical, making them valuable resources for descriptive, theoretical, and applied discussions of language. Because corpus linguistics is a methodology, all linguists even generativists -could in principle use corpora in their studies of language (...) corpora (...) are used for creating dictionaries, studying language change and variation, understanding the process of language acquisition, and improving foreign and second-language instruction. (Meyer 2002, 28)

However, researchers should pay attention to corpus representativeness to be appropriately used as the basis for generalizations concerning a language as a whole; according to McEnery and Wilson (2001), the texts "when taken together, may be considered to 'average out' and provide a reasonably accurate picture of the entire language population in which we are interested" (30). The sample size, method of sampling, and population variability are among the most crucial considerations in assuring representativeness.

3.2.2 Learner corpus research (LCR)

Learner corpora are generally defined as "systematic collections of authentic, continuous and contextualized language use by foreign/ second language learners stored in electronic format" (Callies 2015, 35). They have the same characteristic frequently attributed to corpora, yet, the only difference is that the data in learner corpora come from language learners, that is, from speakers who learn a language that is neither their first language nor has an official function in the country (Granger 2008, 259). Thus, learner corpora represent language produced by foreign or second language learners and are aimed to get an insight into the learners' interlanguage. Furthermore, the data are usually found in an electronic form which makes them more manageable and therefore easier to analyze. In other words, computerized data can be analyzed using several linguistics software tools that are provided for "quick and efficient manipulation of the data via their search" (Granger 2003, 465).

The use of learner corpora as a methodology has increased rapidly since the late 1980s/ early1990s, due to the increased accessibility of computer corpora and software tools (Granger 2008, 259). The digital medium allows for a thorough and detailed analysis of large amounts of learner corpus data and more systematic research of the interlanguage of learners of various language backgrounds. Hence, carrying out research by means of a learner corpus enables "the systematic and (semi-) automatic extraction, visualization and

analysis of large amounts of learner data in a way that was not possible before" (Callies 2015, 35). In addition, access to and analysis of learner corpus data compared to the small scale of data and case studies gives us a more reliable basis to describe and model learner language data (Granger 2003, 456) and allows the generalizability of the results to a broader learner population. As pointed out by Gass & Selinker, "it is difficult to know with any degree of certainty whether the results obtained [from case studies] are applicable only to the one or two learners studied, or whether they are indeed characteristic of a wide range of subjects" (2001, 31). Moreover, learner corpora have contributed to second language acquisition research (SLA) by providing a better description of the interlanguage of learners of various language backgrounds and a better understanding of the factors that influence it, such as L1 transfer.

There are currently several learner corpora that have been compiled (Granger 2008, 261). The most notable learner corpus that covers learners from several L2 backgrounds is the International Corpus of Learner English (ICLE) (Granger/Dagneaux/Meunier 2002; Granger 2020). Learner corpora consisting of various language backgrounds help researchers carry out studies where they can compare non-native speakers to native speakers' language and allow a comparison of different L1 backgrounds to each other, in a so-called contrastive interlanguage analysis (CIA), see figure 3.1.

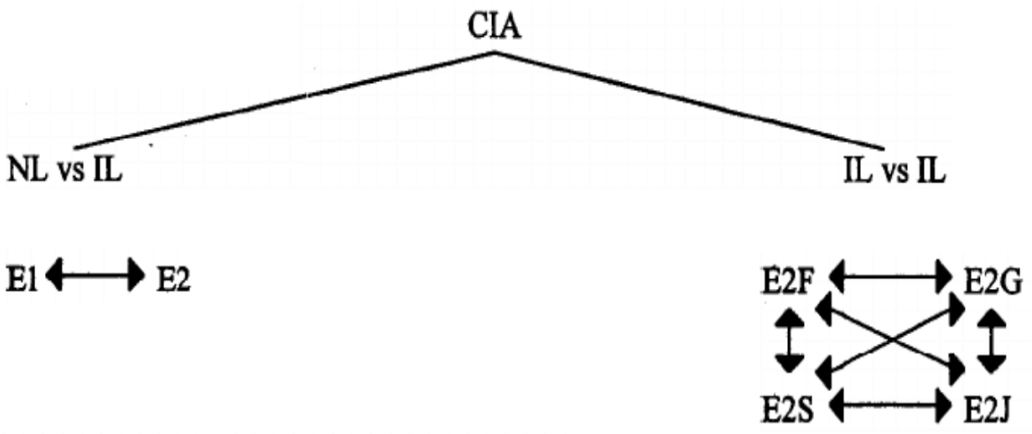


Figure 3.2 1: Contrastive Interlanguage Analysis (Granger 1996) [2]

2. NL = native language, IL = interlanguage, E1 = English as an L1, E2 = English as a foreign language (FL), E2 l = English as a foreign language by speakers with l as L1

The current study belongs to the left-handed side of the CIA, in that Native language will be compared to the interlanguage produced by Norwegian learners of English. Moreover, it can be argues that the two learners populations both of which have Norwegian as their L1, belongs to the right-handed branch, by comparing the interlanguage of different levels of education.

3.3 Material

This section presents the corpora on which this research is based on. To answer the aforementioned research questions (see chapter 1), a comparison will be conducted on the basis of material drawn from three corpora, containing a collection of essays written by intermediate Norwegian learners of English, advanced Norwegian learners, and English native speakers.
The data will be extracted from the following corpora, respectively:

1. The TRAWL corpus (Tracking Written Learner Language).
2. The Norwegian part of the International Corpus of Learner English (ICLE-NO).
3. Louvain Corpus of Native English Essays (LOCNESS).

3.3.1 The TRAWL corpus

The TRAWL corpus is being compiled as a part of the ongoing project Tracking Written Learner Language (Dirdal et al. 2017). The corpus contains authentic texts written by Norwegian pupils as part of their regular school work, which enables research description of the development on their writing skills in their L2. All the texts are collected longitudinally from pupils at Norwegian schools aged 10-19; that is, the data are collected from the same learners over time, allowing for both pseudo-longitudinal and longitudinal studies and allows studies of the development of the pupils' writing skills. The TRAWL writers are younger than both LOCNESS and ICLE-NO writers (see the following sub-section).

Texts are collected from years 5-13 for English and 8-13 for French, German, and Spanish. Moreover, some texts are written in Norwegian to compare L1 and L2 writing development. However, the focus of this study is on the English texts written by

students in 8th and 10th grades (from 2014-2017). In this study, 177 texts will be examined with a total word number of 131,489. The texts in the TRAWL corpus are argumentative essays written by NNS of English, whose first language is Norwegian.

3.3.2 The International Corpus of Learner English (ICLE)

The International Corpus of learner English (ICLE; Granger et al., 2020) was initiated by and coordinated by Sylviane Granger, Université Catholique de Louvain, and it provides a large amount of material for studying advanced learner English. The ICLE corpus is a compilation of mainly argumentative essays written by learners of English whose level goes from higher intermediate to advanced. In the ICLE corpus, the informants are university undergraduate students of English language and literature studies, which makes them "learners who are generally expected to have mastered the basic rules and regulations of the language they are learning" (Lorenz 1999, 10).

The essays have been systematically collected from 25 different mother tongue groups, including Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Turkish, and Tswana, and it contains more than 5.5 million words from 1990 to 2020. The texts focus on various topics which are considered controversial (see list of suggested topics[3]). However, some learner and task variables had to be controlled to ensure comparability across the different corpora, such as learners' proficiency level, age, gender, and country of origin, see figure (3.1).

3. https://uclouvain.be/en/research-institutes/ilc/cecl/corpus-collection-guidelines.html.

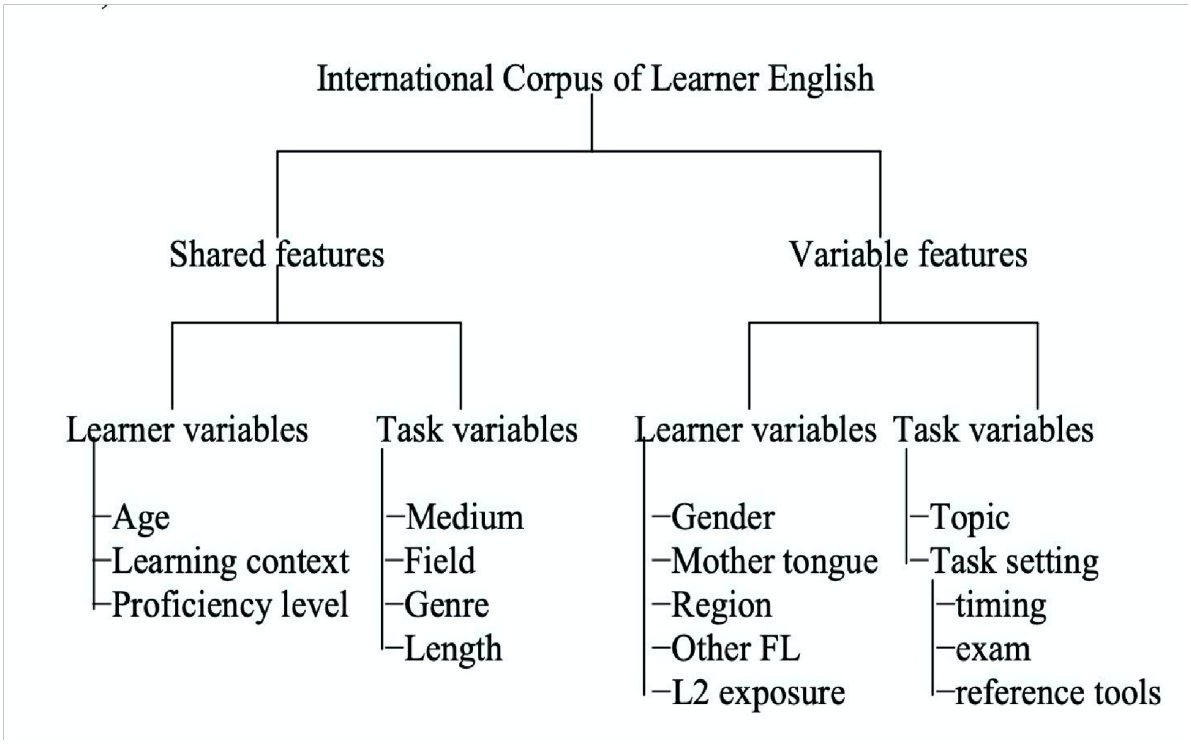


Figure 3.1: ICLE learners and task variables. (Johansson 2008, 115), based on Granger 1998

The Norwegian part of the ICLE (ICLE-NO) corpus was compiled between 1999-2002 by Lynell Chvala and Stig Johansson, University of Oslo. All the contributors are university students who attend different universities/colleges in Norway, and the majority have studied English for less than a year at the university level, indicating a similar level of proficiency (Fossan 2011, 73). However, they vary in features like gender, sex, mother tongue, and other foreign languages (Granger 1998, 9). The Norwegian component consists of more than 300 essays comprising a total of 211,725 words. The average essay length is between 500-700 words.

It should also be stressed that I chose not to use the new ICLE search interface, where the subcorpora have been part-of-speech tagged, because I wanted to make sure that searches in the three corpora were exactly the same, thus yielding comparable results, see section 3.4.1.

3.3.3 Louvain Corpus of Native English Essays (LOCNESS)

The comparable native speaker corpus for ICLE is the Louvain Corpus of Native Essay Writing (LOCNESS), which contains argumentative essays written by British A-level pupils and British and American university students. The texts are, thus, written by novice writers whose first language is English. The essays vary from 400 to 2300

words, with an average text length of 1028 and a total number of words of 324,304. The corpus has the following structur.[4]

- British pupils' A-level essays: 60,209 words
- British university students' essays: 95,695 words
- American university students' essays: 168,400 words

The LOCNESS corpus was specifically designed as a reference corpus for ICLE, as Granger et al. (2009, 42) state, " to ensure comparability with the ICLE data, the Louvain team has collected a corpus of essays written by native English students, the Lovain Corpus of English Essays (LOCNESS), which is the mirror of the ICLE". The writers in ICLE-NO and LOCNESS are nearly the same age and have an almost equivalent educational status, making the two corpora relatively comparable, (see table 3.1).

Granger (1998, 13) points out that the disadvantage of the LOCNESS corpus could be that it is a small corpus that contains material from non-professionals. However, the texts in the LOCNESS corpus are not intended as role models of perfect native writing; instead, they were put together as a reference corpus for ICLE to make the comparison more straightforward.

3.3.4 Comparability of the corpora

Corpus comparability is often an issue in learner corpus research (LCR) as the available corpora have different word numbers and can also differ in text type, register, task setting and genre, (see for example Breeze 2017 on the influence of task variable; Granger and Pequot 2009; and Gentil and Meunier 2018 on register and genre differences). Controlling these variables is essential for the validity of the results.

In terms of comparability of the three corpora in this study, we have seen that the LOCNESS was designed to be comparable to the ICLE corpus. The TRAWL corpus is arguably also comparable to the other two according to many variables outlined in table 3.1. the table shows that the texts in ICLE-NO and TRAWL are argumentative essays written by NNS of

4. See further: http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/CeclProjects/Icle/locness1.htm.

English, whose first language is Norwegian. Still, the learners' proficiency levels and years of study are considerably different.

Table 3.1: Comparability information from NICLE and LOCNESS corpora

Subcorpora	TRAWL	ICLE-NO	LOCNESS
Text language	English	English	English
Number of words	131,489	211,725	324,304
L1	Norwegian	Norwegian	English
Country of oreign	Norway	Norway	US/UK
Genre	Argumentative	Argumentative	Argumentative
Topic	General argumentative topics	General argumentative topics	General argumentative topics
Level of education	Lower secondary school	University students	University students +A-level

Although the corpora differ in size, this will not be a major challenge, as I will work with normalized frequencies.

3.4 Data extraction Method: AntConc

In order to study large quantities of texts and reveal patterns in language, several software programs are available to us, such as WordSmith tools, Lancsbox, and AntConc. The software used in this study to retrieve data from the corpora is AntConc 3.5.8 (Anthony, 2019). It was developed by Dr. Laurence Anthony, a professor at Waseda University, Japan, and is one of the most widely used corpus analysis tools.

The reason for choosing this program is that, as Anthony (2004, 7) notes, "AntConc is a freeware application, making it ideal for individuals." In addition, AntConc is a convenient and user-friendly piece of software that offers robust corpus linguistic tools, including the concordance tool. AntConc has the advantage of having been around for

almost twenty years and has tutorials in 9 different languages (Anthony 2002). Moreover, AntConc can be used for searches dealing with multiple frames. For example, in this thesis, the focus is on the use of the passive voice in academic texts; thus, I was able to search for some of the typical passive voice frames: *was/ are/ is/ am/ were/ be/ being/ been* as well as all the forms of the lemme GET (*get/ gets/ got/ gotten and getting)* followed by words ending in **ed* and **en,* see table 3.2 below.

In sum, the AntConc program is employed in this study to retrieve the passive constructions from the three corpora mentioned in section 3.3.The first step is to find how frequently certain passive types are used by native and non-native speakers of the English language, followed by a second step in which these constructions were used to look for the frequency of *by*-phrase and the lexical association of verbs with the passive voice.

3.4.1 Extraction of data and preliminary analysis

The analytical process in the present research can be summarized in the following steps. The first step was extracting the data from all the corpora mentioned above using AntConc. Thus, a search of each of the eight forms of BE was carried out separately since the corpora are not annotated, i.e., (*is*, *was*, *are*, *were*, *am*, *be*, *being*, *been*) followed by (**ed* or **en*). Moreover, I searched for all the contracted forms for the auxiliary BE (*'s/ 'm/ 're*) and HAVE (*'ve/ 'd*) followed by *been* and (**ed*/**en*), to get as accurate data as possible. See table 3.2.

Table 3.2: Search strings for BE constructions with examples from the ICLE-NO corpus

Forms	**Search strings**	**Example from ICLE-NO**
Present	*is/am /are/'s/ 'm/'re *ed/ *en*	During the practice periods the students **are guided** by a teacher.
Past	*was/were *ed/ *en*	Society **was ruled** by the clergy.
Present/Past Perfective	*has/have/had/'s/'ve/'d been *ed/ *en*	Some of these things **have been mentioned** in the lecture.

Present /Past Progressive	*is/am/are/was/were/'s/'m /'ve being *ed/ *en*	What subject **is being studied**?

According to Quirk et al. (1985, 189), all the passive constructions that lack any aspectual marking are called simple passives. So, the first two types in table 3.2 are titled the present passive and the past passive, respectively. The simple passive consists only of the auxiliary BE or GET followed by the past participle of the lexical verb. Additionally, a medial element can also occur in the verb phrase. However, I searched only for BE/GET+ past participle without intervening words in between because placing an additional element in the verb phrase was problematic as it yielded numerous irrelevant hits.

The perfective and progressive passives (table 3.2) obtain their names from the aspectual markers in the verb phrase, with the pattern (HAVE + *been/got (ten)* + past participle) and (BE + *being/getting* + past participle), respectively.

Although the prototypical periphrastic central passive is the BE-passive, GET-passives will also be included in this study. According to Carter and McCarthy (1999, 47), they are "close to the unmarked passive with BE" i.e., the central periphrastic passive. Moreover, central GET-passives, in line with central BE-passives, have an active equivalent with identical propositional meaning; as Collins (1996, 45) states, they are distinguishable by "their potential relatedness to a propositionally equivalent active clause." In other words, the GET-passives fulfills the criteria in 1-5 below.

Thus, to extract the GET-passive from the corpora, all the forms of GET that signal passive voice "*get*," "*gets*," "*got*," "*gotten*," and "*getting*" followed by the past participle verb were searched for, table 3.3.

Table 3.3: Search strings for GET-constructions with examples from ICLE-NO.

Forms	Search strings	Example form ICLE-NO
Present	*get/gets *ed/en*	When women enter the labour market they tend to **get employe**d as social workes
Past	*got *ed/en*	If someone **got killed** in the Norwegian Saga period, the dead ….
Perfective	*gotten *ed/en* *got *ed/*en*	No hits
Progressive	*getting *ed/*en*	No hits

The data obtained needed further analysis to distinguish central passives from other constructions. Therefore, the second step was to manually examine all the concordance lines in every retrieval result. Next, the relevant results for each corpus were gathered into one .txt-file, and copied and pasted into different Excel documents according to their forms (present, past, perfective, progressive) and type (GET or BE-passive). Once all the examples were obtained, all the instances that clearly were not passive voice were weeded out. Next, I worked through the remaining results and determined the passives that are qualified as central passives. Some basic criteria should be fulfilled to classify the periphrastic passive construction as a central passive. According to Quirk et al. (1985, 167), the central passive:

1. Has purely verbal properties, that is, displays no adjectival characteristics.
2. Has an active counterpart.
3. The auxiliary BE cannot be replaced by a lexical copular verb such as *feel, seem, remain, sound*, and *look*.
4. The subject corresponds to the object in an active version of the clause where the object is the affected entity (patient), and the subject is found as *a by*-phrase (agent) or absent.
5. The past participle cannot be prefixed with the negative *un-* nor modified by adverbial such as *very* and *rather* (since this gives them an adjectival status).

For instance example (3.1) meets Quirk et al. 's criteria for a central passive with an expressed *by*-phrase, hence, it is included in the study:

(3.1) The country **was occupied** by the Europeans for more than 200 years.
(P01013_Y10_CSCC_V0_ORIG.txt 423)

- It displays no adjectival characteristics.
- It has an active counterpart: *the Europeans occupied the country for more than 200 years.*
- The past participle cannot be paired with degree adverbs such as *very and quite*. * *was very occupied*, **was quite occupied.*
- The auxiliary *BE* cannot be replaced by a copular such as *seem* or *feel*. *The country seems/feels occupied by the Europeans for more than 200 years*

All the other passive sentences that did not meet the central passive criteria were discarded, such as semi-passives (3.2) and pseudo-passives (3.3), which are passive in form, but not in meaning. The searches also returned several false hits, e.g. (3.4), where **en* triggered the adverb *often* which is clearly not a passive voice. Furthermore, the V*ing*-form (3.5), non-finite passives (3.6), and modal passive (3.7) were excluded from the analysis due to their infrequent use in the material.

(3.2) Because they **are interested** in what they are doing. (ICLE-NO.txt 06)

(3.3) My stereo, CD's, telephone, jewelry, my husband's firearms, clothes, furniture, everything **was gone** from my home. (LOCNESS. txt 121)

(3.4) There **is often** little room for individuality and alternative perspectives.
(ICLE-NO. txt 0146)

(3.5) Many will object saying that you lose your freedom **being locked up**.
(ICLE-NO. txt 014)

(3.6) A question **to be asked** here is if any of the two opinions express the truth.
(ICLENO.txt 03)

(3.7) He **may have been weeded out** during the first season of play.
(LOCNESS. txt 11)

Separating central passives from other passive types was not straightforward. In some cases, the distinction between verbal and adjectival uses was problematic since the adjectival constructions are identical in form to a past participle. This yielded many borderline cases in which it is difficult to determine whether the past participle is a participial adjective or a verb. Yet, the distinction was made through several different steps according to the criteria mentioned above.

Accordingly, the first step is to make sure that the passive construction is purely verbal and displays no adjectival proprieties. This is tested by adding a degree adverb *very*, *quite*, or *rather* to modify the past participle; if the sentence accepted the modification, it was removed from the list. For example, the adjective *very* can precede the participle *tired* in example (3.8) and *interested* (3.9), while it is not the case in examples (3.10) and (3.11). Accordingly, (3.8-3.9) are not considered central passive, unlike (3.10-3.11).

(3.8) They are driving while they **are tired. (**LOCNESS. txt05)

(3.9) Baseball really needed more fans who **were interested** in the game. (LOCNESS. txt 125)

(3.10) During his imprisonment, Kaliayev **is given** three attempts to save his life. (LOCNESS.Txt 071)

(3.11) Transfer students **were placed** into the desegregated school system. (LOCNESS. txt 112)

Secondly, the presence of the agent in the *by*-phrase supports the verbal interpretation of the verb as in (3.12), and thus, it is considered central passive. Moreover, a passive construction with a potential active counterpart accepting *by*-phrase, as indicated by Quirk et al. (1985), is also classified as central passive. In example (3.13), the agent is unexpressed, but it can be inferred from the context, e.g. (*by the police*), and an active counterpart can be derived: *the police had arrested their father*.

(3.12) During the practice periods the students **are guided** by a teacher. (ICLE-NO)

(3.13) A call that informed them that their father **had been arrested**. (ICLE-NO)

Next, the auxiliary in a central passive construction cannot be replaced by other linking verbs, such as *feel*, *look*, and *seem,* nor can the participle be paired with another adjective. The auxiliary *are* in example (3.14) and *is* in (3.15) are replaceable with the copular verbs: (*women seem/feel/look prepared*) and (*he seems/feels/looks determined*), respectively. Besides, a true adjective can be added to the participles: (*are prepared and ready*) and (*he is determined and insistent*). Thus, they are eliminated from my list.

(3.14) The women **are prepared** for the job. (LOCNESS. txt 0110)

(3.15) He **is determined** to carry through with his train of thought. (LOCNESS. txt 01504)

Moreover, the past participles cannot be preceded by the negative prefix *un-*. The examples above (3.14) and (3.15) accept the prefix *un-*, *unprepared*, *undetermined*, and hence, are considered to be adjectival rather than verbal. While the prefix *un-* cannot be applied to the central passives, as in e.g. (3.10) and (3.11) provided above.

Finally, from a semantic point of view, the participle central passive is assumed to express an action carried out on the subject of the sentence or describe an activity rather than describing the resultant state of the subject. On the other hand, the stative passive focuses on the result, not on the action itself, and they, as Biber et al. describe them, "are like constructions with copular BE plus adjective" (1999, 940). Quirk et el. (1985) call them "adjectival" or "pseudo-passive." Consider the following examples:

(3.16) a. First of all their friend **is killed** by a vicious sailor. (LOCNESS. txt 148)

b. Instead, he chooses to fight Claudius, and Hamlet **is killed**. (LOCNESS. txt 170)

(3.17) a. Indian men, women and children **was executed** by soldiers. (LOCNESS.Txt19)

b. He **was executed** on May 20, 1982. (LOCNESS. txt 14)

Examples (3.16a) and (3.17a) are classified as central passive since, among other things, the participles do not accept adverb modification, nor can the auxiliary BE be replaced by a lexical copular. Furthermore, these sentences have an active counterpart clause where the subject corresponds to the object in an active version of the clause, as follows:

- **A vicious sailor** killed their friend.
- **Soldiers** executed Indian men, women and children.

If we compare with the (b) examples, the agentive *by*-phrases in examples (3.16b) and (3.17b) are absent, and thus, more analysis is needed to qualify them as central passives. In (3.16b), the sentence can be extended to add an agent, which is already mentioned earlier in the context (*Hamlet is killed by Claudius*); thus, it is classified as central passive.

In example (3.17b), the actor is unexpressed and has two different readings, dynamic and stative. The verb (*to execute*) is a transitive verb that requires an object to fulfill the meaning. In this sense, the participle denotes an event, and an agent can be added; hence, it has a dynamic interpretation (*he was executed by someone*). In the case of a stative (adjectival) interpretation (*he was executed* (*not alive*) *at that time*), the sentence denotes a description of a state of being that cannot be extended to add an agent and, therefore, does not have an active counterpart. Because of their ambiguity, these types of sentences are removed from the list.

When it comes to the combination of passive voice with the *by*-preposition phrase, the extraction of long passive constructions unsurprisingly yielded some irrelevant results. Thus, I had to go through every *by*-phrase, to remove all the instances where the preposition *by* was not followed by the doer, such as when *by* is considered an instrument or locative. In the examples below (3.18-3.20), *by* is not followed by the performer of the replacement or punishment, but rather by the instruments which are (*computers and robots*) in (3.18) and *games* in (3.19), and by the location (*prison*) in (3.20).

(3.18) People **are replaced by** computers and robots. (ICLE-NO. txt 09)

(3.19) Fantasy developing activities **are replaced by games**. (ICLE-NO. txt 08)

(3.20) Norwegian people **are punished** for their illegal actions **by prison**. (ICLE-NO.txt 023)

The third step, of a quantitative nature, is normalizing the raw frequencies to instances per 100,000 words. Normalized frequency can tell us how common the passive is relative to the total number of words in each specific corpus and allows us to compare corpora from different sizes directly and more accurately (Biber et al. 1998, 263f.). Therefore, the results were normalized by dividing the raw frequencies by the corpus size from which

they were extracted and then multiplying them by 100,000. Later, the normalized frequencies were rounded to the nearest whole number. A quantitative overview of the data will be given in chapter 4, followed by a qualitative analysis of the central passive in the material.

4 ANALYSIS AND FINDINGS

4.1 Introduction

The analysis presented in this chapter employs the methods outlined in the previous chapter. The analysis and findings follow the order of the research questions presented in section 1.2. This chapter is divided into three main sections. Section 4.2 starts with a quantitative analysis of the data to firstly give an overview of the overall frequency of the passive voice across the three corpora, the frequency of the GET and BE-passive, and the frequency of short and long passives. Then, in the qualitative section 4.3, some potential factors will be put forward to help account for the choice of one type of passive over the other and to discuss potential contributing factors that may affect the choice of BE and GET-passives and short and long passives. Finally, the last section 4.4 contains a list of the most frequent verb forms occurring in the BE and GET-passive. Next, the verbs will be listed in two tables to compare and discuss semantically the past participles associated with each passive auxiliary. In other words, I will argue if the most frequent verbs associated with BE and GET share semantic characteristics, for example, if they are dynamic or indicate negative connotations.

4.2 Quantitative analysis

In this chapter, frequency is presented in different ways. Raw frequency is the actual number of occurrences in the corpora and is provided in table columns headed (Raw Frequency of passive). Since the corpora investigated vary in size, the normalized frequency per 100,000 words is given in the columns headed (Frequency per 100K words). Normalized frequency reports frequency against a common base of normalization as a proportion of each corpus, (see e.g McEnery and Hardie 2012, 49).

4.2.1 Frequency of the passive voice across the corpora

The first objective of this study was to investigate if Norwegian speakers of English produce passive constructions in English as frequently as native speakers do. This section is thus dedicated to counting and determining the proportion of all the passive forms in the three corpora: LOCNESS, ICLE-NO, and TRAWL. The count of passive voice constructions

includes all central passive verb phrases, with or without the *by*-phrase, as marked by the presence of the lemmas BE and GET.

To begin with, we will first look at the overall distribution of the passive voice in the three selected corpora (table 4.1) to determine how often each student group tends to employ passive constructions in their writing.

Table 4.1: Frequency of the central passive voice in LOCNESS, ICLE-NO and TRAWL

	No. of words in corpus	**Frequency of passive (raw)**	**Frequency (per 100Kwords)**
LOCNESS	324,304	769	237.12
ICLE-NO	211,725	377	178.06
TRAWL	131,489	155	117.88

The comparison between the three corpora decodes differences between the native speaker corpus and the other two NNS corpora. For example, the number of passive forms in the LOCNESS corpus is 237.12 per 100,000 words, compared to 178.06 in ICLE-NO and 117,88 per 100,000 words in TRAWL. The numbers indicate that the L1 writers in LOCNESS deploy the passive voice more frequently than the undergraduate university students in ICLE-NO, though the long exposure of the L2 learning and use, and much more than the young TRAWL learners. Thus, the data support our first hypothesis in section 2.5, that "the frequency of the passive voice in the writing of Norwegian learners will be lower than in native speaker writings." Besides, the fact that the passive voice is almost twice as frequent in LOCNESS compared to TRAWL, implies that Norwegian pupils from lower secondary schools tend to avoid the passive and markedly prefer the active voice.

The comparison between the Norwegian pupils in TRAWL and the Norwegian undergraduate university students in ICLE-NO shows that the passive voice occurs more frequently in the writing of those who are older and at a more advanced level. Thus, not unexpectedly, the use of the passive voice increases with age and level of proficiency. Moreover, learners acquire new and more sophisticated forms of passives at different

learning stages. This fact is reflected in the corpora investigated; for example, only one instance of the progressive passive was found in the TRAWL corpus (0.65%), compared to 21 (5.5 %) in ICLE-NO and 39 (5.1%) in LOCNESS.

The results from our data confirm what Hinkle (2004), Gilquin (2008), and Granger (2013) report in their studies. They all observed consistent underuse of the passive voice among the learner groups compared to native language speakers (see section 2.5). The reason, among other things, could be that Norwegian students are advised to avoid using the passive in their (formal) writing when the active voice is possible since the active voice is usually more direct and vigorous than the passive (Strunk and White 1979,18). In other words, it could be teaching induced; as Hinkle (2004) outlines, the passive voice is not properly addressed to a certain degree because of the traditional separation of teaching grammar that covers the use of voice and teaching writing.

Another reason could be traced to the fact that the English language lacks the morphological passive in contrast to the Norwegian language. Thus, this may have influenced the Norwegian students to select the easiest way to express their idea by replacing the passive voice with its active counterpart.

4.2.2 Frequency of BE and GET-passives

When looking at the different passive voice types, the total number of passives mentioned above is realized by BE and GET-passives together. Therefore, in this section, we will examine the total number of each of them separately. Tables 4.2 and 4.3 below offer an overview of BE and GET-passives' frequency. In addition, the number of BE and GET-passive constructions in comparison to the total number of central passives in the corpus is provided in the ratio columns.

Table 4.2: Frequencies of BE-passives in LOCNESS, ICLE-NO and TRAWL

	No. of words.	**No. of passives**	**Frequency of BE-passives (raw)**	**Frequency of BE-passive (per 100k words)**	**Ratio%**
LOCNESS	324,304	769	750	231.26	97.53%

ICLE-NO	211,725	377	350	165.31	92.83 %
TRAWL	131,489	155	126	95.83	81.29%

Table 4.3: Frequencies of GET- passive in LOCNESS, ICLE-NO and TRAWL

	No. of words	**No. of Passives**	**Frequency of GET-passive (raw)**	**Frequency of GET-passive (per 100k words)**	**Ratio (%)**
LOCNESS	324,304	769	19	5.86	2.47%
ICLE-NO	211,725	377	27	12.75	7.16%
TRAWL	131,489	155	29	22.05	18.71%

The results of the present quantitative analysis suggest that the BE-passive occurs considerably more frequently than the GET-passive in all three corpora. This implies that Norwegian speakers of English at different proficiency levels prefer BE-passive to GET-passives in their course writing, the same way native speakers do. Thus, the BE-passive is the unmarked variant by Norwegian learners and native speakers of English, whereas the GET-passive is the marked one. These findings are in agreement with the second hypothesis that "BE-passives will be significantly more frequent than GET-passive in Norwegian writings," see section 2.5.

However, it is interesting to see, as demonstrated in tables 4.1 and 4.2 above, the LOCNESS corpus has the highest frequency of passive voice overall and the highest proportion of passives with BE. The BE-passives are the most frequent in LOCNESS, with 231.26 per 100,000 words, and the least frequent in TRAWL, with 95.83 occurances per 100,000 words. When it comes to GET-passives, the LOCNESS corpus contains a comparatively low frequency of 5.86 per 100,000 words, followed by the ICLE-NO corpus of 12.27 per 100,000 words and then the TRAWL corpus with 22,05 per 100,000 words.

Concerning the passive ratio, 97.5% of all passive occurrences in the LOCNESS corpus are BE-passives, leaving only 2.5% for GET-passive constructions. In ICLE-NO corpus, the BE-passives account for 92.83% of all the passives, compared to 7.16 % of GET-passives. The highest percentage of GET-passive usage is found in the TRAWL corpus, accounting for 18.71 %, while the BE-passive has a percentage of 81.29 %. These numbers are visualised in figure 4.1.

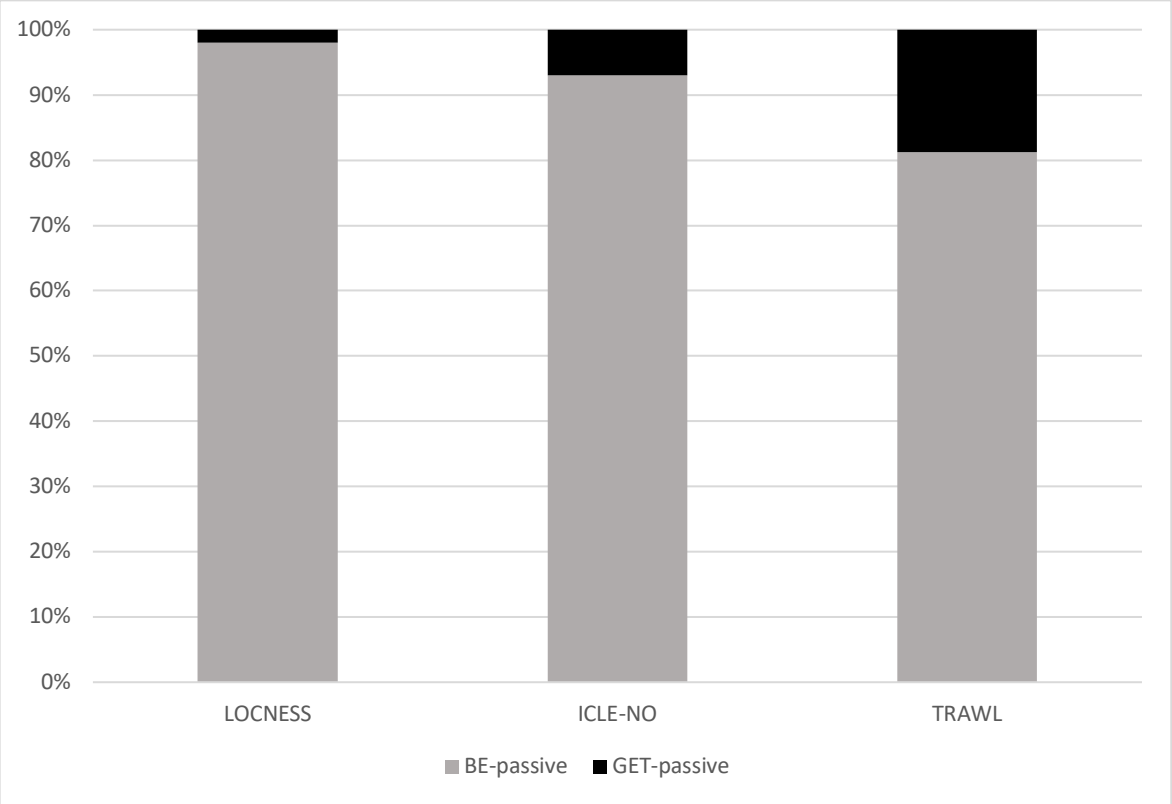


Figure 4.1: Ratio of BE-passives and GET-passives in the three corpora

Accordingly, one can conclude that, relatively speaking, Norwegian pupils at the secondary school level employ the GET-passive more frequently in their written assignments compared to Norwegian undergraduate university students and L1 native speakers. The same conclusion is reflected in Marchman et al.'s (1991) study, which reported that children are more likely to use the GET-passive. An interpretation could be that TRAWL students seem to transfer the GET-passive from their spoken, informal domain to their writing or even from the Norwegian spoken language by directly translating the *BLI*-passive into the GET-passive. In other words, it could be a case of L1 transfer.

The nature of the topics (i.e. the subject matter) involved in each corpus can also play an essential role regarding the frequency of the BE-passive and the GET-passive. The BE-passive is believed to be "the passive of being," and GET-passive is "the passive of becoming" (Jespersen1909,49b). As stated earlier (2.3.2) and in (4.3.2) below, the GET-passive is associated with dynamic verbs denoting an action in which the agent somehow affects the action. On the other hand, the BE-passive can also be used both with dynamic and non-dynamic verbs. Thus, when students want to indicate the agent's responsibility for the event, they tend to use the GET-passive, see example (4.1-4.2), and the BE-passive is the first choice when their own point of view is discussed (4.3-4.4), as in the narrative register. The differences between the two alternative passive constructions will be explored in the following sections.

(4.1) In these countries the criminals either **get placed** in prison or they have to do.... (ICLE-NO.txt044)

(4.2) They **get whipped** if they don't do their work. (TRAWL- P01013_Y08_TAUS_V0_ORIG. txt301)

(4.3) There are two main "subjects" that often **are discussed**, censorship according to sexual exploitation and …. (ICLE-NO.txt 0100)

(4.4) All these challenges **have been mentioned** a lot of times.

(TRAWL P01013_Y10_CSCC_V0_ORIG. txt421)

However, it is not straightforward to draw any general conclusions at this stage, as there are considerable proficiency level and age differences between the three groups, and more evidence is needed to draw an objective conclusion.

4.2.3 Frequency of short and long passives

One of the main reasons to choose the passive voice over the active one is to avoid mentioning the agent; hence, it is expected that most of the instances of the passive voice are expressed without an agentive *by*-phrase. According to Quirk et al. (1985, 164f.), only one out of five passive sentences in English has an expressed agent. Similar findings are reported in Svartvik's (1966) analysis of corpus data: four out of every five English passive sentences are short passives (cf. section 2.2). Similarly, Biber et al. (1999, 938) report a relation

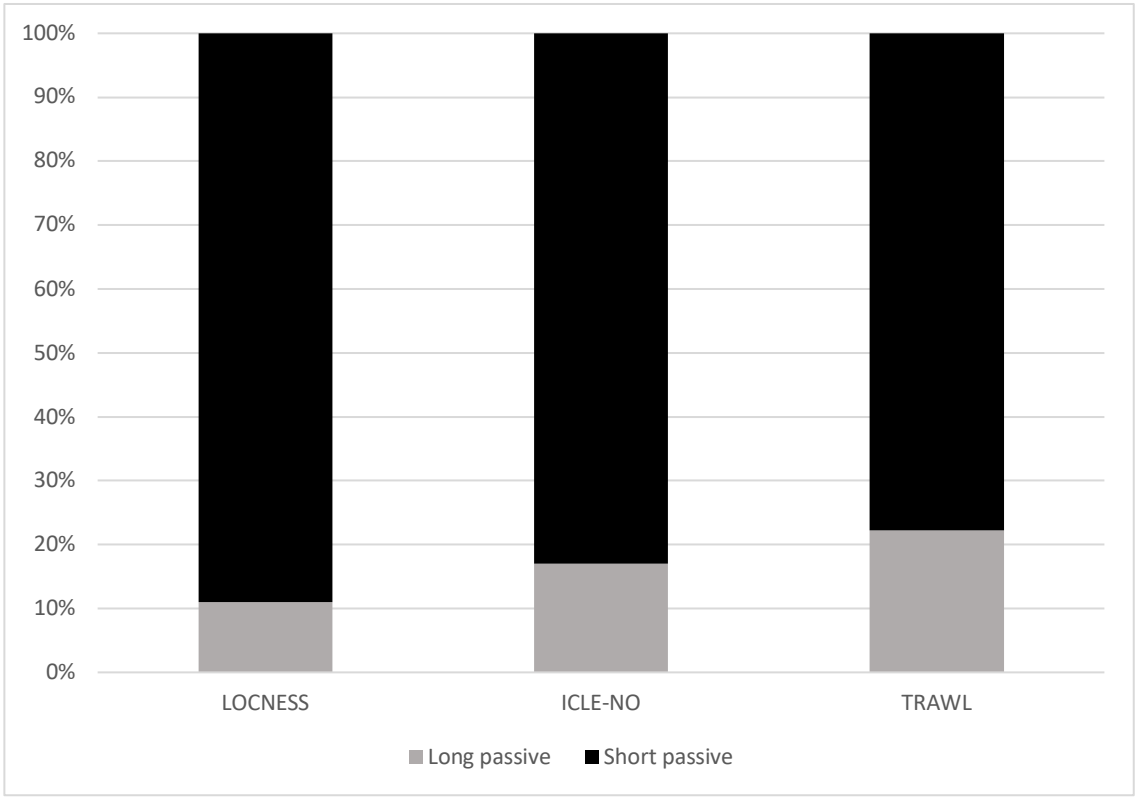
between one agentive passive to seven non-agentive ones in fiction and academic discourse. As already stated in the chapter 2, passive constructions that contain an agentive *by*-phrase are often referred to as 'long' or 'agentive passive,' in contrast to 'short' or 'agentless' passive where the agent is omitted (Biber et al. 1999, 935). The complement of the preposition *by* in the agentive *by*-phrase typically corresponds to the subject or the actor in the active counterpart of the sentence.

The results of the present quantitive part of the thesis confirm that the overwhelming number of passive constructions have an unexpressed agent (short passives). Table 4.4 shows the frequency of short and long passives across the corpora.

Table 4.4: Frequency of short passives and long passives across the corpora

		Long passive		**Short passive**	
	Total No. of passive	No. of long passives (raw)	Ratio%	No. of short passives (raw)	Ratio%
LOCNESS	769	83	10.79%	686	89.21%
ICLE-NO	377	63	16.71%	314	83.29%
TRAWL	155	35	22.51%	120	77.41%

In the LOCNESS corpus, short passives (89.21%) are around eight times as frequent as long passives (10.79%). In ICLE-NO, out of 377 examples, 314 (83.29%) do not have an expressed agent, while only 63 (16.71%) are with an expressed agent, and they are close to the ratio of one to five, as Quirk et al. (1985, 164f) stated. Finally, the TRAWL corpus has the highest ratio of long passives (22.51%) compared to (77.41%) short passives. Thus, short passives are predominantly more frequent than long passives in all three corpora. This implies that the style of non-native Norwegian students, to some extent, have resembled that of English native writer in academic writing; see figure 4.2 for a more convenient comparison. This again confirms one of the passive voice's main purposes which is to deliberately leave the agent unspecified.



Fiqure 4.2: Ratio of short passives and long passives across the corpora

We will now compare the proportion of long BE-passive and short BE-passives to that of long and short GET-passives. As seen in tables 4.5 and 4.6, the proportion of short and long BE-passives compared to short and long GET-passives is quite similar in the three corpora. i.e., the long BE and GET-passives occur less frequently than short BE and GET-passives in all three corpora; hence, the short form is preponderant in its use. This was suggested by Huddleston and Pullum (2005, 243), and it is supported by Collins (1996), who reports from the analysis of the Lancaster Oslo/Bergen (LOB) Corpus, the Brown Corpus, ACE, and CIE, that 92% of GET-passives did not include a *by*-phrase. Similar findings are reflected in the present study.

Table 4.5: Frequency of long and short passives associated with BE

BE -passives

Corpus	Long passives		Short passives	
	Raw frequency	Ratio%	Raw Frequency	Ratio%
LOCNESS	81	10.8%	669	89.2%
ICLE-NO	61	17.43%	289	82.57%
TRAWL	32	25.40%	94	74.60%

Table 4.6: Frequency (proportion) of long and short passives associated with the GET

GET-passive

Corpus	Long passives		Short passives	
	Raw Frequency	Ratio%	Raw frequency	Ratio%
LOCNESS	2	11.76%	17	89.47%
ICLE-NO	2	7.41%	25	92.60%
TRAWL	3	10.34%	26	89.66%

In the LOCNESS corpus, short forms of BE and GET-passives are over eight times as frequent as their long forms; in other words, 89.2% of BE-passive sentences and 89.47% of GET-passive sentences occurred with unexpressed agents, compared to 10.8% and 11.76%

with an expressed agent, respectively. In the ICLE-NO, the overwhelming majority of GET-passives (92.60%) occur without an agent compared to 82.57% of BE-passive constructions. Finally, the agentive phrase is expressed only in 3 cases (10.34%) of the GET-passives clauses in the TRAWL corpus and in 32 clauses (25.40%) of BE-passives constructions.

Though the proportion of short and long BE-passives is similar to the proportion of short and long GET-passives in the three corpora, it is quite clear that GET-passives are even more likely to occur without an agent. Proportionally, the short GET-passive occurs more frequently compared to the long GET-passive in ICLE-NO and TRAWL than the short BE-passive, compared to the long BE-passive, while the proportion is almost the same in LOCNESS. In addition to this quantitative contrast, there are also other differences between the two (short and long) alternative passive constructions, which will be discussed in the following sections.

4.3 Qualitative analysis.

This section presents a more in-depth analysis to investigate the factors that may govern the choice of one type of passive over the other. In the first part (section 4.3.1), I will give an overview of factors that are said to influence the choice of GET-passives and BE-passives. The second part (4.3.2) is devoted to studying the factors that affect the choice of the short passive and long passive; in other words, when the agentive *by*-phrase is expressed and when it is omitted. Both sections will discuss the presence of these factors in the current material.

4.3.1 Factors influencing the use of the BE-passive and the GET-passive

Based on previous research, I outlined (in chapter 2) some of the main differences between the BE-passive and GET-passive. Three factors are investigated here: verb dynamicity, adversativity, and the Subject's animacy and responsibility. In this section, a qualitative analysis is performed to examine to what extent these factors seem to influence the students' choice in all three corpora.

4.3.1.1 Verb dynamicity

To reiterate some of the observations made in chapter 2, Huddleston and Pullum (2002, 1442) state that GET-passives are found only with dynamic verbs; that is, verbs that denote an action. Quirk et al. (1985, 162) comment on the advantage of using GET-passives only with dynamic verbs, as this helps to avoid stative/dynamic ambiguity, "[t]he GET-passive provides a convenient way of avoiding the passive with BE in cases where there is a potential confusion between the normal passive interpretation and that of the statal passive." On the other hand, Sasaki (1999) states that GET-passives focus on the described action but may sometimes communicate a "statal passive" in line with the BEpassive that can occur both with dynamic and statal verbs. To find out if similar tendencies are observed in this study, the lexical verbs associated with BE and GET in each corpus will be investigated and sorted according to their type: dynamic and non-dynamic verbs.

Dynamic passives are identified by putting their active version in the progressive, i.e., if the verbs accept the progressive, like in (4.5-4.8), they are considered dynamic:

(4.5) When women enter the labour market they tend to **get employed** as social workes. (Somebody are employing them). (ICLE-NO.txt 011)

(4.6) Where thousands of soldiers **get killed** at the border every day. (Somebody are killing them). (TRAWL P01018_Y10_WAPE_V0_ORIG. txt1032)

(4.7) My little business **got destroyed** by the army. (The army are destroying them). (TRAWL P01026_Y10_CSCC_V0_ORIG. txt521)

(4.8) Every time a boxer **gets punched** in the head his brain moves vigorously. (Something is punching them). (LOCNESS.tx t02)

All GET-passives (100%) in the three corpora are associated only with dynamic verbs, which is to be expected. Based on these findings, it can be said, as reported by Quirk et al. (1985, 162), that GET-passives seem to be restricted to dynamic verbs that denote an action. That is to say, both native speakers and Norwegian speakers of English tend to employ the GET-passive only with dynamic events. Thus, the dynamic property of the main verb plays a vital role in the choice of the GET-passive. Here are more examples of dynamic verbs associated with the GET-passive from the corpora:

(4.9) But in all cases the criminals **get placed** in a cell.
(ICLE-NO. txt 045)

(4.10) We often see in movies that prisoners **get raped** by their cellmates.
(ICLE-NO.txt 05)

(4.11) They **get whipped** if they don't do their work.
(TRAWL P01013_Y08_TAUS_V0_ORIG.txt 301)

(4.12) When students drop out or **get kicked out**.
(LOCNESS. Txt 016)

All the GET-passives in examples (4.9-4.12) are followed by dynamic verbs *place, rape, whip,* and *kick out.* Furthermore, it is worth mentioning that all the verbs in the examples above convey a negative connotation; this factor will be dealt with in the next section.

When it comes to the BE-passive, the choice does not exclusively depend on the dynamic properties of the main verb; thus, BE-passives can be followed by dynamic verbs (4.13-4.15) and nondynamic verbs (4.16-4.19). Consider the following examples from the LOCNESS corpus:

(4.13) 35% of homicide victims **are killed** by someone they know.

(4.14) He **is beaten** almost to death.

(4.15) I believe such works **were written** for other reasons.

(4.16) Should himself be murdered to fulfill what **is considered** by many to be justice.

(4.17) This kind of thoroughness **is expected** by the audience, and the writers.

(4.18) It **is regarded** by many people as being traditional.

(4.19) ….an actual social problem that **is seen** by everyone.

Examples (4.13-4.15) involve BE+ main verb *kill, beat,* and *write* to convey an activity. While the verbs in examples (4.16-4.19) *consider, expect, regard,* and *see* are not considered dynamic but non-dynamic verbs since they do not represent clear-cut dynamic activity. So, the fact that BE-passive can be associated with both dynamic and non-dynamic verbs gives us an indication of why BE-passives occur more frequently than GET-passives in all three corpora. Thus, the dynamic propriety of the verb is one of the factors that may influence or restrict the students when choosing GET-passives. However, the preference for the

BEpassives over the GET-passives is not only determined by the nature of the verbal group but there are also other grammatical components that are key factors in the choice of the GETpassive; these are discussed below.

4.3.1.2 Adversativity

Another critical case where the GET-passive tends to be more restricted in use than the BEpassive is when the main verbs refer to negative or adversative circumstances. As Biber et al. (1999, 481) observe, the GET-passives typically appear with verbs that have "negative connotations, conveying that the action of the verb is difficult or to the disadvantage of the subject." It has also been proposed by Carter and McCarthy (1999) that the GET-passives are predominantly used for expressing a negative speaker stance. However, Leech et al. (2009,156f) observe that the GET-passive can be adversative and non-adversative. Adversative is when the GET-passive triggers negative connotations and non-adversative indicates something neutral or positive.

In this section, I will investigate if GET-passive only co-selects verbs that refer to negative actions or if it can also occur with positive or neutral verbs, as the BE-passive does. The aim is to find out whether adversativity is one of the common factors that may influence native speakers of English and Norwegian learners of English to select GET-passives in their writing. To give a clearer picture, the ratio of the adversative and non-adversative GETpassive is given in table 4.7.

Table 4.7: Frequency of the adversative and non-adversative GET-passives

	Adversative GET- passive		**Non-adversitive GET- passive**	
Corpus	Raw frequency	Ratio (%)	Raw frequency	Ratio (%)
LOCNESS	17	89.47 %	2	10.52%
ICLE-NO	26	96.30%	1	3.70%
TRAWL	27	93.10%	2	6.90%

In 89.47% of cases, the GET-passives in the LOCNESS corpus are adversative, while 10.52% are non-adversative. On the other hand, in the ICLE-NO corpus, 96.30% of the GET-passives are adversative vs. 3.70% non-adversative. Furthermore, the adversative GET-passive is also higher than the non-adversative GET-passive in TRAWL: 93.10% vs. 6.90%, respectively. Thus, the LOCNESS has the highest proportion of the non-adversative GET-passive while the ICLE-NO has the lowest. However, the frequency are very low here, so these proportions are hard to determine with any certainty.

The characteristic used to identify the verb as adversative is whether the GET-passives refer to "a state of affairs that is signaled contextually by the conversational participants as unfortunate, undesirable, or at least problematic" (Hatcher 1949, 441). In other words, those verbs that refer to undesirable events perceived as unfavorable for the subject are classified as adversative verbs. To determine the adversativity of the verbs, I followed a test for adversativity developed by Persson (1990, 52), i.e., if the answer to the question "is it worse to be X than not to be X?" was "Yes," then the passive was labeled "adversative," as in the following examples (4.20-4.22); otherwise, the passive is labeled non-adversative.

(4.20) Suddenly the girl **gets kidnapped**, and the boy have to rescue her. (TRAWL P01018_Y10_ARWO_V0_ORIG.txt 211)

(4.21) Thousands of soldiers **get killed** at the border every day. (TRAWL P01018_Y10_WAPE_V0_ORIG.txt 1032)

(4.22) Once in a while some of them do **get arrested**. (ICLE-NO.txt 09)

Examples (4.20-4.22) illustrate the use of adversative sense since the verbs *kidnap, kill,* and *arrest* express an adversative meaning that conveys a negative effect on the subjects *the girl, thousands of soldiers* and *some of them*, respectively.

However, there are also some cases in the corpora where the GET-passive occurs with verbs that express a more neutral sense. In these cases, the negative output can be inferred from the context, i.e., the verb is neutral while the context is adversative. Here, the sentence is classified as adversative, as in the examples below:

(4.23) I **got treated** like a dog many times. (TRAWL P01026_Y10_CSCC_V0_ORIG.txt522)

(4.24) It is much too easy to **get carried away** in aggresive feelings.
(LOCNESS.txt 024)

On the other hand, the BE-passive arises in our data with adversative and non-adversative verbs and in adversative and non-adversative contexts. This is illustrated in table 4.8.

Table 4.8: Frequency of the adversative and non-adversative BE-passive

	Adversative BE- passive		**Non-adversative BE- passive**		
Corpus	Raw frequency	Ratio (%)	Raw frequency	Ratio (%)	Total (%)
LOCNESS	107	14.37%	643	85.73%	750 (100%)
ICLE-NO	48	13.71%	302	86.29%	350 (100%)
TRAWL	17	13.49%	109	86.51%	126 (100%)

Table (4.8) displays the proportion of adversative and non-adversative verbs associated with the BE-passive. Compared to the GET-passive, the BE-passive is used more frequently in a non-adversative sense. In other words, the BE-passive shows a much higher proportion of positive/neutral occurrences than constructions that convey the negative sense in all three corpora. Comparing the NS with NNS, the proportion of adversative and non-adversative passives are very similar, as the ratio proves.

Based on this data, one can draw a tentative conclusion that the GET-passive is predominantly used with adversative verbs in an adversative context. At the same time, the BE-passive is not as restricted to the verb's adversity. Accordingly, another reason why the BE-passives occurred more frequently than the GET-passives in the corpora is that the BE-passive can more typically be associated with verbs conveying both an adversative and a non-adversative sense. In contrast, GET-passives are typically confined to verbs involving adversity.

4.3.1.3 Subject's responsibility and animacy

The choice between BE and GET-passives depends not only on the properties of the main verb (dynamicity and adversativity) but also on the other grammatical constituents in the sentence. One possible factor that influences the use of one type over the other is the degree of responsibility and animacy attributed to the Subject, i.e., the Subject's responsibility and animacy. This section will discuss the influence of the Subject referent on the choice of GETpassive and BE-passive by NS and non-native Norwegian learners.

A. Subject's responsibility

There is agreement that the GET-passive focuses more on the degree of responsibility for the action involved on the part of the subject. Quirk et al. (1985, 161) note that the GET-passive "puts the emphasis on the subject rather than the agent, and on what happens to the subject as a result of the event." According to Huddleston (1984, 445), "GET lends itself more readily than be to the imputation to the subject-referent of some measure of initiative or responsibility." Further to the above, Huddleston and Pullum (2002, 1442) report that the "subject referent is seen as having an agentive role in the situation, or at least having some responsibility for it."

Hence, the GET-passive is mainly used when the subject-referent (patient) has responsibility for the event or influences the situation (Lakoff 1971; Sussex 1982; Collins 1996). On the other hand, the sentence is formed with BE when the agent (actor) takes, or is given, responsibility for the situation described and retains some control. Similar features are noted by Lasnik and Fiengo (1974), who suggest that when a passive sentence is formed with GET, patient control is assumed; when the same sentence is formed with BE, agent control is assumed. This is typically exemplified in examples (4.25) and (4.26):

(4.25) If someone **got killed** in the Norwegian sage period…. (ICLE-NO.txt. 06)

(4.26) Their friend **is killed** by a vicious sailor. (LOCNESS. txt 148)

In (4.25), the Subject of the GET-passive *someone* is in some way involved or partly to blame in the action of *killing*. In comparison no such inference can be drawn about the Subject of the BE-passive in example (4.26), rather the agent *a vicious sailor* has the control.

B. Subject's animacy

A central feature of the English passive voice is its ability to occur with animate and inanimate Subjects. However, it has been reported that an overwhelming proportion of the passive voice occurs with inanimate entities since it corresponds to the direct Object in its active counterpart, which has the participant role of Patient and is typically inanimate. (Svartvik 1966, 50f, Teleman et al. 1999, 4:390, Van Nice and Dietrich 2003, 829)

Looking at the passive types separately, it has been proposed that GET-passives are more likely to feature an animate Subject[5], whereas BE-passives can take both animate and inanimate Subjects (Teleman et al. 1999, 4:390). Concerning the GET-passive, Sasaki (1999) state that human subjects are most frequently found with the GET-passive, making it more semantically flexible than the BE-passive. Hence, this sub-section examines the proportions of animate and inanimate Subjects in the three corpora (Tables 4.9, 4.10, 4.11). Then, it looks at the differences between the two passive types to examine the animacy effect on BE/GET-passives' production and to find out if the same tendencies arise from our data.

Table 4.9: Subject's animacy in LOCNESS

Subject animacy	**BE-passive**		**GET-passive**	
	Raw frequenc	Ratio (%)	Raw frequenc	Ratio (%)
Animate	214	28.53%	15	78.95%
Inanimate	536	71.47%	4	21.05%
Total	750	100%	19	100%

5. Humans, animals, personal pronouns, and any noun that refers to a group of people or citizens, such as *the society*, were regarded as animate Subjects

Table 4.10: Subject's animacy in ICLE-NO

Subject animacy	**BE-passive**		**GET-passive**	
	Raw frequency	Ratio (%)	Raw frequency	Ratio (%)
Animate	152	43.42%	27	100%
Inanimate	198	65.57%	0	0%
Total	350	100%	27	100%

Table 4.11: Subject's animacy in TRAWL

Subject animacy	**BE-passive**		**GET-passive**	
	Raw frequency	Ratio (%)	Raw frequency	Ratio (%)
Animate	25	19.84%	22	75.86%
Inanimate	101	80.16%	7	24.14%
Total	126	100%	29	100%

Tables 4.9-4.11 show the proportion of animate and inanimate Subjects of BE and GET-passive constructions in the LOCNESS, ICLE-NO, and TRAWL corpora, respectively. Based on the tables above, there is an overall tendency for the BE-passive Subject to be inanimate (71.47%, 65.57%, 80.16%), whereas the GET-passive Subject is more likely to be animate (78.95%, 100%, 75.86%).

The intergroup analysis shows that the BE-passive takes both animate and inanimate Subjects in the three corpora. However, there is a clear predominance of inanimate Subjects by NS in LOCNESS and NNS in TRAWL. In line with Hinkle (2002), who observes that inanimate nouns in English are commonly used as Subjects. In the ICLE-NO corpora, although the percentage of inanimate Subjects is higher than the animate ones, the gap is smaller, representing 43.42% and 65.57% respectively.

As regards the Subject in GET-passive construction, the picture is inverted. Animate Subjects prevail over inanimate Subjects in all three corpora. The percentage of animate and inanimate Subjects in LOCNESS and TRAWL corpora are nearly the same, representing 78.95%, 75.86% for animate Subjects, and 21.05%- 24.14% for inanimate Subjects, respectively. Interestingly, all the GET-passive constructions produced by NNS in ICLE-NO have animate Subjects (100%).

Regarding the characteristics attributed to the Subject, the animacy of the Subject referent in GET-passive shows that most animate subjects are human subjects with active participation and responsibility for the action described in the clause. That would explain why the inanimate Subject is infrequent with GET-passive; i.e., the inanimate Subject cannot normally be assigned the notion of responsibility. Nevertheless, that does not mean the GET-passive does not take inanimate Subjects. On the contrary, GET-passive can take inanimate subjects, as the tables above show. Givón and Yang explain this (1994, 120-121) by stating:

> When the subject of the GET-passive is inanimate, thus itself incapable of responsibility, some human associated with the subject, or with the event in some capacity, may either retain responsibility, be involved in the action or be adversely affected by the result. (Givón and Yang 1994, 120-121)

Thus, the inanimate non-human subjects in GET-passive construction are not responsible for the action but rather the affected entity, as example (4.27) from the TRAWL corpus displays. The subject *videos and pictures* are not responsible for the action in the clause *being shared and liked*, but they are affected by the result.

(4.27) Videos and pictures **gets shared and liked** every day. (TRAWL)

On the other hand, the BE-passive puts the focus more on the agent than the Subject. Thus, both animate and inanimate Subjects frequently occur in the Subject slot. Consider the following examples from ICLE-NO:

(4.28) I believe that this statement **is used**, unfairly, by people.
(4.29) During the practice periods the students **are guided** by a teacher.

The focus in both sentences is on the animate human agents *people*, and *teachers* rather than on the inanimate Subject *this statement* and the animate Subject *the students*, respectively.

According to the given-new principle, given information is related to background information and can occur anywhere in the sentence, excluding the final position. New information is related to the focused information that is normally placed at the end of the sentence, after the verbs, (cf. Šimík, Wierzba and Kamali 2014). Consequently, as long as the new or focused information is the agent, the Subject or given information in BE-passives can take both an animate or inanimate Subject that presents an objective account of the action or has passive involvement in the event in terms of responsibility and control.

4.3.2 Factors influencing the use of the short and long passive

As already noted in earlier sections, most English passive sentences occur without an agent (agentless). Proportionally, the GET-passives are more likely than the BE-passive to occur without an agent, as they tend to leave the agent unexpressed to emphasize the patient and event (cf. Carter and McCarthy1999, 44). In addition, according to linguists such as Dušková (1994) and Huddleston and Pullum (2002), there is a tendency to leave the agent unexpressed in cases where the agent is a general human actor, unknown or the speaker intends to avoid identifying the agent. Thus, agentless passives highlight the process in question rather than the doer or the performer of the action.

However, there are some cases where we can hardly omit the agent from the passive voice sentences, as it displays crucial information and is semantically indispensable. In this section, we will investigate those instances by giving examples for each factor to say something about the characteristic of the expressed and unexpressed agents and to find out what influences students to use one type or the other.

A. Factors influencing the use of short passive

As we have seen, constructing a sentence in the passive voice is a central strategy to avoid mentioning or specifying the agent; therefore, most passive clauses do not contain the agentive *by*-phrase. As mentioned in section 4.3.1.3, GET-passives focus more on the Subject of the sentence to the extent that the agent is completely dropped. At the same time,

BE-passives tend to focus equally between the patient and the agent. In this sub-section, we will look at different agentless examples from the corpora to give an account of the cases where the agent is left out.

With a view to identifying the reasons for dropping the agent from the passive sentences, Stanley (1975) notes that there is often no clear-cut way to determine whether an agent is deleted consciously or not, yet, the motivation is often intellectual laziness. Quirk et al. (1985, 161) and Carter and McCarthy (1999, 52) suggest that the basis for leaving the agent phrase unexpressed is the low information value the agent has. Mihailovic (1965, 6) points out that the deletion of the *by*-phrase is not always one of personal preference. He goes on to state that the verb of the sentence determines the choice of suppressed or expressed agent; some verbs enable agent deletion, e.g., *broke*, *signed*, and *faked*; some never permit agent deletion, e.g., *possessed*, *actuated*, and *succeeded.* Further, Huddleston and Pullum (2002, 1446) report that short passives can be used to avoid identifying the person responsible for the situation.

In table 4.12, I have listed several cases to illustrate the factors that influence the students to use short passive. However, I should highlight that there were a few cases of doubt and uncertainty in the process of classifying the material. For example, it was not clear-cut in some cases if the authors' motivation was not to identify the agent or if it was unknown to them. These instances are listed in table 4.12 under the head (unspecified). Nevertheless, the classification made in this section is based on the context in which the passive clauses are used. Table 4.12 below displays the five most frequent factors that make students in the three corpora omit the agentive phrase from the clause.

Table 4.12: Proportion of the factors that influence the use of the short passive

Corpus	LOCNESS	ICLE-NO	TRAWL	Total
Recoverable from the context	223 (32.51%)	90 (28.66%)	22 (18.33%)	**335 (29.86%)**
General human actor	290 (42.27%)	160 (50.95%)	45 (37.5%)	**495 (44.19%)**
Unknown or unimportant	40 (5.83%)	25 (7.96%)	6 (5%)	**71 (6.34%)**
Authorial Passive	33 (4.81%)	20 (6.36%)	10 (8.33%)	**63 (5.63)**
Avoid identifying the agent	6 (0.87%)	5 (1.59%)	5 (4.16%)	**16 (1.43%)**
unspecified	94 (13.70%)	14 (4.46%)	32 (26.66%)	**140 (12.5%)**
Total	**686 (100%)**	**314 (100%)**	**120 (100%)**	**1120 (100%)**

We can infer from the table 4.12 that the agentive *by*-phrase is frequently omitted, firstly, when the agent of the sentence is assumed to be recoverable from the context, and as Huddleston states, "this type can be described formally by the familiar agent-deletion transformation, which suppresses *by* + NP provided the latter is a pro-form" (1971, 104), consider the following example:

(4.30) We often see in movies that prisoners **get raped** by their cellmates or even sometimes **get killed**. (ICLE-NO. txt 05)

(4.31) Should boxing be banned? Every time a boxer **gets punched** in the head his brain moves vigorously. (LOCNESS.txt 02)

In example 4.30, the ICLE-NO student is reporting a series of actions *raping* and *killing* carried out by the same agent *cellmates*. Thus, to avoid repetition, the agent is mentioned in the first clause; prisoners *get raped by their cellmates* while it is left out in the second one *get killed.* In example (4.31), the agent can be elicited from the context, and the sentence would be, i.e., *get punched by other boxers*.

Another large class of instances that have unexpressed agents is when the agent is a general human actor that can be replaced by *one* or *people*, accounting for 44.19% of all the short passive occurrences in the three corpora. For instance, the active version of example (4.32) would be *people commit a large part of crime under the influence of alcohol or drugs*. The agent in example (4.33) can be *by someone,* and the active version would be *someone paints pictures,* and *someone writes and publishes books*.

(4.32) A large part of crime **is committed** under the influence of alcohol or drugs. (ICLE-NO.txt 0184)

(4.33) Pictures **are painted**, and books **written** and **published**. (ICLE-NO.txt 045)

Sometimes the agent is unknown or unimportant in the sentence, as in examples (4.34) and (4.35). For example, the ones who named the place (4.34) and wrote the sticky note (4.35) are unknown to the writer and even with no interest to the hearer to understand the sentence. What matters are the name of the place and the telephone number of the Russian hackers.

(4.34) He finds a paper with a place that **is named** jalapeno city. (TRAWL P01010_Y08_TAUS_V0_ORIG.txt 271)

(4.35) The telephone number of the Russian hackers **was written** on a sticky-note. (TRAWL P01026_Y08_TAUS_V0_ORIG.txt 391)

Furthermore, a relatively large number of passives occur without an agent when the actor is the author, i.e., the author of a book or article. This type of passive is called authorial passive, and it is used to maintain the objective flavor of the text (Dušková 1994, 260) without reference to the writer. In example (4.36), the author intended not to mention the first person *I*, making their personality invisible.

(4.36) Another reason for ceasing the practice of euthanasia as it **is stated** previously is the consequences America may suffer because of euthanasia. LOCNESS.txt 0432

In other cases, the short passive is used because the speaker wants to avoid identifying the agent responsible for the action (Huddleston and Pullum 2002, 1446), as examples (4.37-4.39).

(4.37) That's also why people sometimes **are punished** for things we didn't even know. (ICLE-NO.txt 017)

(4.38) We don't want a war where thousands of soldiers **get killed** at the border every day. (TRAWL P01018_Y10_WAPE_V0_ORIG.txt 1032)

(4.39) However, Candide **is kicked out** of the castle in Westphalia. (LOCNESS.txt 0614)

B. Factors influencing the use of long passive

The passive with an expressed agent (long passive) occurs significantly less frequently in our data than the passive with an unexpressed agent (short passive). That follows from the fact that the main function of the passive is to leave the agent unexpressed deliberately (see chapter (2) and section (4.2.3) for more material). However, Shintani (1979) sheds light on some cases where the agentive *by*-phrase tends to be overtly expressed to explicitly mention who performed the action. These cases are illustrated in table (4.12) with the frequency of each factor from the corpora.

Table 4.13: The proportion of the factors that influence the use of long passive.

Corpus	Proper name	Indefinit noun phrase	Unexpected Inanimate agent	Other (definite noun phrase)	Total
LOCNESS	20 (24.09%)	17 (20.48%)	15 (18.07%)	31 (37.35%)	**83 (100%)**
ICLE-NO	2 (3.17%)	21 (33.33%)	32 (50.79%)	8 (12.69%)	**63 (100%)**

TRAWL	12 (34.29%)	8 (22.86%)	5 (14.28%)	10 (28.57%)	**35 (100%)**
TOTAL	**34 (18.78%)**	**46 (25.41%)**	**52 (28.72%)**	**49 (27.07%)**	**181 (100%)**

According to Shintani (1979), the Agent in passive construction tends to be expressed in the following cases:

Firstly, when the agent is a proper name indicating an artist, an inventor, a discoverer, or an innovator:

(4.40) The book **was written** by "Roald Dahl". (TRAWL P01010_Y08_BAMA_V0_ORIG.txt 71)

(4.41) This philosophy **was founded** by John Locke of England. (LOCNESS.txt 0137)

The agents *Roald Dahl* and *John Locke of England* in examples (4.40)-(4.41) are expressed to indicate the proper name of the writer of the book and the founder of the philosophy, respectively. The agents are expressed and placed at the end of the sentence as the most important information in the clause (information structure).

This category occurs most frequently in the writing of the TRAWL students (34.29%), followed by LOCNESS (24.09%) and ICLE-NO (3.17%). Thus, the use of the long passive, in this case, differs markedly between the Norwegian students at advanced education level, the TRAWL learners, and LOCNESS native speakers. As mentioned earlier, the essay topics in the three corpora are more or less dissimilar, despite being of the same genre. This can explain why, for example, the proper name agent is not as frequent in ICLE-NO as in TRAWL.

Secondly, when the agent is an indefinite noun phrase conveying new information. Table 4.13 above demonstrates that the Norwegian students in ICLE-NO have the highest percentage (33.33%) in this category compared to the LOCNESS (20.48%) and TRAWL corpora (22.86%).

(4.42) During the practice periods the students **are guided** by a teacher. (ICLENO.txt 075)

(4.43) The country of Illyria in Central Europe **is ruled** by a Fascist Regent. (LOCNESS.txt 0194)

(4.44) The majority of the murders in SC **are committed** by friends, relatives, neighbors, …. (LOCNESS.txt 0373)

The agentive phrases in example (4.42)-(4.44) consist of the preposition *by* followed by an indefinite noun phrase (NP) *a teacher*, *a Fascist Regent*, and*friends,relatives, neighbors*, respectively, to highlight the new information in the clause.

Finally, the agent is expressed when the agent is an unexpected inanimate noun.

(4.45) Hunting is a horrible sport as the fox **is hunted down** by a pack of dogs. (LOCNESS.txt 020)

(4.46) Pangloss **is ravaged** by syphilis contracted from Paquette. (LOCNESS.txt 0279)

The agents in the above examples (4.45)-(4.46) are inanimate nouns that are unexpected for some reason. The way the fox dies, in example (4.45) and Pangloss's disease (4.46), were somehow unexpected. Compared to the other categories, this group is the most frequent one in ICLE-NO (50.79%) and the least frequent in LOCNESS (18.07%) and TRAWL (14.28%). It seems like NS in LOCNESS and NNS in TRAWL similarly use the long passive, while the ICLE-NO students have other tendencies.

In addition, the data in our corpora reveals another large category in which the *by*-phrase is expressed, i.e., when the agent is a definite noun phrase, accounting for 27.07% in total; see example (4.47)-(4.48) below. In 37.35% of long passive occurrences in the LOCNESS corpus, students tend to express the agent when it is a definite NP to stress more emphasis, making it the largest category in this corpus. The TRAWL students express the definite NP agent in 28.57% of all the long passive occurrences compared to 12.69% in ICLE-NO.

(4.47) Because the cars **are abused** by the riders.
(LOCNESS.txt 0335)

(4.48) My little business **got destroyed** by the army.
(TRAWL P01026_Y10_CSCC_V0_ORIG.txt 521)

Accordingly, compared with students in LOCNESS and TRAWL, the ICLE-NO students are more likely to leave out the agent when it is a definite noun phrase and express it when it is conveying new information as when the agent is an indefinite noun phrase or unexpected Inanimate agent.

4.4 The most frequent verbs to occur with BE and GET-passives in the corpora

The last part of this chapter is intended to look further into the most frequent verbs that occur in the passive voice in the three corpora. The aim is firstly to identify the verbs most frequently co-selected with the BE-passive compared to those that occur with GET-passive in each corpus. Secondly, to find out if NS and NNS behave similarly in terms of lexical choice and to examine if the verbs in the same group share common characteristics.

The five most frequent passivized verbs associated with the BE-passive are elicited manually and listed in table 4.14. Then, the ratio for each verb is counted by dividing the raw frequency of the verb in passive voice by the number of BE-passives in the corpus. The same is also done for verbs associated with the GET-passive in table 4.15.

Table 4.14: The five most frequent verbs with the BE-passive in the corpora

	LOCNESS		**ICLE-NO**				**TRAWL**	
verbs	Raw frequency	Ratio %	verbs	Raw frequency	Ratio %	verbs	Raw frequency	Ratio%
Give	43	5.73 %	Give	17	4.85%	Write	11	8.73%
Use	41	5.46%	Use	15	4.28%	Call	7	5.55%
Take	15	2%	Force	14	4%	Treat	6	4.76%
Kill	10	1.33%	Create	10	2.85%	Publish	4	3,17%
Argue	9	1.20%	Call	8	2.28%	Name	4	3.17%
Total	118	15.72%		64	18.26%		32	20,66%

By comparing the native speakers of English in LOCNESS to NNS in ICLE-NO, the result shows that the top two verbs *give* and *use* in LOCNESS and ICLE-NO are the same, and their proportion does not differ considerably. Thus, the verb *give* exhibits a strong association with the BE-passive, followed by BE+ *used* both in LOCNESS and ICLE-NO. However, more differences are witnessed further down the list, where in the two corpora the last three verbs differ, and their proportion in the passive is also very different. E.g., the verbs *force*, *create* and *call* with a ratio of (4%, 2.85, 2.28%) in ICLE-NO occur twice as frequently as the verbs *take*, *kill* and *argue* (2%, 1.33%, 1.20%) in LOCNESS, respectively. Hence, more verbs are repeated more often in ICLE-NO than in LOCNESS. This can also be witnessed by looking at the total ratio number in each corpus, i.e., in ICLE-NO, the five most frequent verbs, mentioned in table (4.14) make up 18.26% of all BE-passive instances in the corpus, compared to 15.72% In LOCNESS.

On the other hand, the top five verbs found in the TRAWL corpus differ completely from those in LOCNESS and ICLE-NO. None of them has any representation among the most frequent verbs in the other two groups. For example, while the most frequent verb, *give* represents 5.73% in LOCNESS and 4.85% in ICLE-NO, the most frequent verb in TRAWL *write* has a ratio of 8.73%; which is the highest of all verbs in the table. Moreover, TRAWL's total passive ratio is strikingly higher than in LOCNESS and ICLE-NO, representing 20.66%.

As hinted at, the total passive ratio reveals the lexical variation in students' writing. In LOCNESS, the five most frequent verbs account for 15.72% of all verbs associated with the BE-passive, giving room to 85% of other different verbs. In ICLE- NO, the five verbs represent 18.26% leaving 81.74% for other verbs. The highest passive ratio in the top five, with 20.66%, is in the TRAWL corpus. Thus, the NNSs in TRAWL have the least lexical variation, followed by the ICLE-NO and then the LOCNESS. This is expected, as younger and less proficient learners tend to clings to their lexical teddy bears (Hasselgren 1994).

Regarding the semantic characteristic of the verbs mentioned, the table 4.14 above contains adversative and non-adversative verbs. For example, verbs such as *give*, *use*, *write*, and *publish* are non-adversative, whereas *kill* and *force* are considered to be adversative/negative verbs that convey negative effects on the subjects. As noted in section 4.3.1.2, one of the main factors influencing choosing BE-passive over GET-passive is its ability to appear with

both adversative and non-adversative verbs. Moreover, all the verbs in table 4.14 across the three corpora are dynamic verbs denoting an action.

When it comes to the GET-passive, the aim was to obtain the five most frequent verbs that occur in the GET-passive, but since the number of GET-passives is much lower than BE-passives, many verbs occurred only once, which makes it challenging to include all those verbs in the table. Thus, only the verbs that occurred more than once are investigated.

Table 4.15: The three most frequent verbs co-associated with the GET-passive in the corpora

LOCNESS			**ICLE- NO**			**TRAWL**		
Verbs	Raw frequency	Ratio (%)	Verbs	Raw frequency	Ratio (%)	Verbs	Raw frequency	Ratio (%)
Punch	2	10.52%	Kill	3	11.11%	Tease	5	17.24%
Sentence	2	10.52%	Place	2	7.40%	Bully	3	10,34%
Total	4	21.04%		5	18.51%		8	27.58%

As shown in table 4.15, only a few verbs are used more than one with the GET-passive. The most frequent GET-passive in the LOCNESS corpus is GET+ *punch*/*sentence*; each represents 10.52% of all GET-passive instances. GET+ *kill* accounts for 11.11% in ICLE-NO, and GET+ *tease* for 17.24% in TRAWL. The total passive ratio in ICLE-NO (18.51%) is lower than the passive ratio in TRAWL (27.58%) and LOCNESS (21.04%). That, once again, confirms that the students in the TRAWL corpus have the least lexical variation compared to students in ICLE-NO and LOCNESS.

Concerning the type of the lexical verbs, the results obtained show that the GET-passive is restricted to dynamic and adversative verbs, that is, verbs that denote action and not its outcome and convey that the action is to the disadvantage of the subject. All the verbs in table (4.15) are dynamic verbs representing a dynamic activity and have negative connotations, except the verb *place*, which is a neutral verb. Yet, the negative output can be inferred from the context of both occurrences.

Summarizing this chapter, the following should be noted. Firstly, the BE-passive and the GET-passive are found in the writing of both NS and NNS of English. However, the frequency of the passive occurrences in the Norwegian learner corpora is lower than in the native-speaker corpus. Comparing L2 English learners at different learning stages, the Norwegian undergraduate students in ICLE-NO employ the passive voice more frequently than Norwegian pupils in TRAWL. Yet, the three groups have the BE-passive as the unmarked variant, while the GET-passive is the marked one. Secondly, some factors are mentioned above that would influence the use of GET- and BE-passives. One of the most important of these is dynamicity, i.e., the GET-passive is only associated with dynamic verbs, whereas the BEpassive is found with dynamic and non-dynamic verbs in the corpora.

Furthermore, the short passives occur almost eight times more frequently than the long passive in the corpora investigated, and a very similar trend was noted. The qualitative analysis revealed that the prominent cases where the agentive *by-phrase* is left unexpressed are when the agent is a general human actor or recoverable from the context. However, in other instances, such as when the agent is an indefinite noun phrase, the agent is usually expressed.

Eventually, investigating the most frequent passive verbs revealed something about the lexical variation in the corpora. For example, the top five passive verbs in LOCNESS account for 15.72% of all the BE-passive occurrences,compared to 20.66% in TRAWL. Thus, native speakers in LOCNESS have more variation in their lexical choice than TRAWL students.

5 CONCLUSION

5.1 introduction

The following provides a brief overview of the work contained in this thesis. The present study has investigated the usage of English passive constructions in texts produced by Norwegian NNS of English and NS of English. In chapter 2, the review of some background literature shed light on the similarities and differences regarding passive constructions in Norwegian and English. The aim is to determine how the Norwegian learners of English at different age and learning stages, and Native speakers of English use the English passive voice. Later, chapter 2 revealed the complexity of finding a clear-cut definition for the passive voice; as a result, the term passive gradient has emerged. The investigation in this study is based on the Quirk et al.'s (1985) classification scheme and is limited to the central BE and GET-passives, as specified within this passive gradient.

In order to collect our data, three corpora were identified and examined in chapter 3. Two of them contain texts written by Norwegian students from different learning stages (ICLE-NO, TRAWL), and one containing texts written by native speakers of English (LOCNESS). The primary purpose of this study was to explore the overall frequency of the passive voice in the Norwegian NNs corpora and the NS corpus and to compare the usage of passive voice types and their lexical associations between the three groups.

By reviewing the literature, one can conclude that while there have been plenty of studies examining the frequency and the use of passive voice in the production of L1 and L2, there has not been much research on the production of the passive voice by learners from the same L1 but at a different level of education.

5.2 Summary of the major findings

The results of the quantitative and qualitative analysis have provided answers to the research questions that were posed to help fulfill the aim of the study, these are once again listed below with a summary of the findings.

Research question1: Do Norwegian learners of English use the passive voice to the same extent as native speakers do?

This research question was investigated by examining differences in passive voice frequency between ICLE-NO and TRAWL writers, and English native speakers in LOCNESS. The normalized frequencies presented in chapter 4, section 4.2.1 strongly suggest that the undergraduate university students in ICLE-NO and the young TRAWL learners significantly underuse the passive voice compared to NS of English. For example, the passive voice is almost twice as frequent in LOCNESS compared to TRAWL. In addition, a difference in passive frequency was also witnessed among the Norwegian NNS groups. The passive voice occurs more frequently in university students' writing in ICLE-NO compared to Norwegian pupils in TRAWL. Thus, the results point to the fact that Norwegian students have a tendency to the avoid passive voice in (academic texts). The explanations could be that Norwegian learners were advised to avoid using the passive voice in the writing register or that the lack of the morphological passive in English forced Norwegian learners to replace it with its active counterparts. Moreover, the findings confirm that the use of the passive voice increases with age and proficiency.

1A: How frequent is the use of the central BE-passive vs. the central GET-passive?

To answer this question, the frequency of BE-passive and GET-passive were counted separately in each corpus. Not surprisingly, the results revealed that the BE-passive was indeed more frequently attested than the GET-passive by both native speakers of English and non-native Norwegian learners, confirming that the BE-passive is dominant in the written register (academic prose). The highest proportion of BE-passives was found in LOCNESS (97.53%), followed by ICLE-NO (92.83%) and then the TRAWL corpora (81.29%).

The picture is reversed when it comes to the GET-passive. The TRAWL corpus had the highest GET-passive proportion (18.71%), followed by the ICLE-NO (7.16%) and then the LOCNESS (2.47%). While Native speakers mainly produce BE-passive in the formal discourse, Norwegian learners of English have more variation between the GET and BE-passives. Taking into account the differences between English and Norwegian noted in the literature review and confirmed in this study; we conclude that L1 influence and genre transfer could be key factors here.

1B: **How frequent is the use of the long vs. the short passive?**

One of the main purposes of using the passive voice is to avoid mentioning the agent. The results confirmed this claim and showed that the proportion of short passive is much higher than the long passives in written English in all three corpora. Native speakers in LOCNESS had the highest ratio of the short central passive; that is, the agent was left unexpressed in 89.21% of the passive sentences compared to 10.79% with an overt subject. Norwegian learners of English in the ICLE-NO corpus dropped the agent in 83.29%, followed by the TRAWL corpus with a ratio of 77.41%.

A further distinction that was investigated was between the long BE and GET-passives vs. the short BE and GET-passives. The data revealed that GET-passives are more likely to occur without an agentive *by*-phrase than the BE-passives in ICLE-No and TRAWL corpora. For example, in the TRAWL corpus, 25.40% of the BE-passive instances occurred with an agent, while only 10.34% of GET-passive instances had the agent expressed. On the other hand, in LOCNESS, the proportion of short (89.2%) and long (10.8%) BE-passive were similar to that of the GET-passive with 89.47% and 11.76%, respectively. Thus, native speakers of English drop the agent in the BE-passive more frequently than non-native speakers, which explains why the total short passive in LOCNESS was the highest compared to the other two corpora.

Research question 2: What factors influence the students to choose between BE and GET-passives, on the one hand, and between short and long passives, on the other?

The distribution of the passive voice is subjected to many factors. Based on the qualitative part of this study, several factors have been mentioned that may affect the proportion of BE-passives and GET-passives, i.e., verb dynamicity, adversativity, subject's responsibility, and animacy. The NS and the Norwegian learners produced GET-passives only when the verb is dynamic; at the same time, the BE-passive was found with dynamic and non-dynamic verbs. Moreover, dynamic verbs associated with the GET-passive express negative connotations, in contrast to the BE-passive, which can also be associated with verbs that convey positive and

neutral connotations. Hence, GET-passives are restricted to dynamic adversative verbs, whereas the BE-passive is not so.

Regarding the factors that influence the use of the short passive, students in the three corpora omit the agentive *by*-phrase when the agent is a general human actor. This category had the highest proportion in LOCNESS (42.27%), ICLE-NO (50.95%), and TRAWL (37.5%), with a total ratio of 44.19%. On the other hand, the agent in the passive constructions was expressed in many other cases. In LOCNESS, native speakers most frequently had the agent overt when the agent was a definite noun phrase (37.35%). Norwegian learners of English in ICLE-NO expressed the agent in 50.79% of all the long passive construction when the agent is inanimate. When it comes to the TRAWL pupils, the highest proportion of the long passive is when the agent is a proper noun (24.09%). The results also point in the direction of overuse from the ICLE-NO writers. The ICLE-NO students express the agent more than the other two groups in cases where the agent conveys new information. As a result, the ICLE-NO corpus has the highest proportion of the definite noun phrase agent and unexpected Inanimate agent but the least proportion of definite noun phrase agent.

Research question 3: Do Norwegian learners of English and native speakers have similar lexical variety when producing the passive voice?

The data revealed that the BE-passive frequently occurs with many recurrent verbs, in contrast to the GET-passive, which is only common with few verbs. The top five verbs with the BE-passive had the lowest overall frequency in LOCNESS, accounting for 15.72%, followed by ICLE-NO (18.26%), and then TRAWL (20.66%). The two most common verbs with the GET-passive have a proportion of 18.51% in ICLE-NO, 21.04% LOCNESS, and 27.67% in TRAWL. Thus, Norwegian learners in TRAWL have the least lexical variation of all. The younger and less proficient generation in TRAWL seems to clings more to their lexical teddy bears than the other two groups. Thus, the lexical variation increases with the age and proficiency level.

5.3 Limitations and further studies

The analysis carried out in this study has several limitations that must be addressed. Firstly, the study does not account for all the types of passives; it is confined to the central BE and GET-passives. Thus, HAVE-passives, pseudo-BE-passives, and semi-BE-passives were not part of this study. Thus, the other passive types should be investigated further to get a sounder picture of the use of the passive voice and to uncover if the current results still hold for all types of passives.

Secondly, the process of classifying the material was sometime difficult. For example, the eliciting of the central passive from the data was far from clear-cut, as it was hard to decide whether the participle had adjectival or verbal properties. In addition, the classification of verbs as dynamic or non-dynamic left several cases of doubt. Thus, this research has left many borderline cases unexplored and needs to be investigated further.

Regarding the corpora investigated, I should emphasize that although the three corpora contain essays of a similar genre, the essay topics are dissimilar; hence, one would expect a potential effect related to the results obtained. A study where NS an NNS perform the exact same writing task with the same task instructions and in similar task settings would be important to confirm the findings of this study. Furthermore, this study examines the passive voice only in the written register. Thus, a future study that compares the use of the passive voice in the written and the spoken registers would be necessary.

Moreover, while all the texts in LOCNESS and ICLE-NO are investigated, only the texts written by the 8th and 10th grades (2014-2017) are examined in the TRAWL corpus. Therefore, expanding the material to include all the texts in TRAWL would be interesting. Besides, a longitudinal study tracking the use of the passive voice among TRAWL learners in different grades would be helpful in further studies to reveal something about the developmental factors and proficiency level.

Finally, the results of this study raise some interesting research questions for further analysis. The present study has made it evident that further study into the passive voice in a reference corpus containing Norwegian essays written by Norwegian learners or translated texts is needed to understand the cross-linguistic influence and L1 transfer better.

References

Alexiadou, A. 2005. A note on non-canonical passives: the case of the get-passive. In *Organizing Grammar* (pp. 13–21). De Gruyter Mouton.

Anthony, L. 2002. AntConc (Version 1.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software.

Anthony, L. 2004. "AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit". An Interactive Workshop on Language e-Learning, 2004. IWLeL. 7-13.

Anthony, L. 2019. AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software.

Baker, M. 1992. *In Other Words: A Coursebook on Translation*. Routledge.

Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University.

Biber, D, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar Spoken & Written English Cased*. London, England: Longman.

Callies, M. 2015. "Learner Corpus Methodology." In *Cambridge Handbook of Learner Corpus Research*, edited by Sylviane Granger Gaëtanelle Gilquin Meunier, 35–56. Cambridge University Press.

Carter, R, and Michael McCarthy. 1999. "The English Get-Passive in Spoken Discourse: Description and Implications for an Interpersonal Grammar." *English Language and Linguistics* 3 (1): 41–58.

Carter, R., & McCarthy, M. (2006). Cambridge Grammar of English. Cambridge: Cambridge University Press.

Coetzee, John M. 1980. "The Rhetoric of the Passive in English." *Linguistics* 18 (3– 4). https://doi.org/10.1515/ling.1980.18.3-4.199.

Collins, P. C. 1996. Get-passives in English. *World Englishes*, *15*(1), 43–56. https://doi.org/10.1111/j.1467-971x.1996.tb00091.x.

Crawford, J. L. 2012. "Developmental Perspectives on the Acquisition of the Passive." Storrs, USA: University of Connecticut.

Ellis, R, and Gary Barkhuizen. 2005. *Analysing Learner Language*. London, England: Oxford University Press.

Engdahl, E. 2006. “Semantic and syntactic patterns in Swedish passives”. In *Demoting the Agent. Passive, middle and other voice phenomena*, B. Lyngfelt and T. Solstad (eds), 21–45. Amsterdam and Philadelphia: John Benjamins.

Faarlund, J. T, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Folse, K. 2009. *Keys to Teaching Grammar to English Language Learners: A Practical Handbook*. Ann Arbor: University of Michigan Press.

Fossan, H. 2011. *The writer and the Reader in Norwegian Advanced Learners' Written English: A corpus-based study of writer/reader visibility features in texts by Norwegian learners of English and native speakers of English* (Master thesis).

Gao, C. Z, Marianne Celce-Murcia, and Diane Larsen-Freeman. 1999. “The Grammar Book: An ESL/EFL Teacher's Course.” *TESOL Quarterly* 34 (2): 372.

Gass, S. M., & Selinker, L. 2001. *Second language acquisition: An introductory course* (2nd ed.). Mahwah, N.J: LEA.

Gilquin, G. 2008. Combining contrastive and interlanguage analysis to apprehend transfer: detection, explanation, evaluation. In *Linking up contrastive and learner corpus research* (pp. 1–33). BRILL.

Gilquin, G, and Sylviane Granger. 2021. “The Passive and the Lexis-Grammar Interface: An Inter-Varietal Perspective.” In *Perspectives on the L2 Phrasicon: The View from Learner Corpora*, edited by Sylviane Granger, 72–98. Channel View Publications. https://doi.org/10.21832/9781788924863.

Givón, T., & Yang, L. 1994. The rise of the English GET-passive. In *Typological Studies in Language*. John Benjamins Publishing Company.

Granger, S. 1983. *The Be + Past Participle Construction in Spoken English*. New York: Oxford.

Granger, S. (1996). *From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora*.

Granger, S. 1998. 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:

Granger, S. 2003. “Error-Tagged Learner Corpora and CALL: A Promising Synergy.” *CALICO Journal* 20, no. 3: 465–80. http://www.jstor.org/stable/24157525.

Granger S. 2008. Learner corpora. In Lüdeling, A. & Kytö, M. (eds.) *Corpus Linguistics. An International Handbook. Volume 1*. Berlin & New York: Walter de Gruyter, 259275.

Granger, S. 2013. “The Passive in Learner English Corpus Insights and Implications for Pedagogical Grammar.” In *Learner Corpus Studies in Asia and the World. School of Languages and Communication*, 5–15. Ishikawa; Kobe University; Kobe.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. International Journal of Learner Corpus Research, 1(1), 7–24. https://doi.org/10.1075/ijlcr.1.1.01gra.

Granger, S, Joseph Hung, and Stephanie Petch-Tyson, eds. 2002. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam, Netherlands: John Benjamins Publishing.

Granger, G., Dagneaux, E., Meunier, F. & Paquot, M. (2009*). International Corpus of Learner English v2*. Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, S., Dupont, M., Meunier, F., Naets, H. & Paquot, M. (2020) The International Corpus of Learner English. Version 3. Louvainla-Neuve: Presses universitaires de Louvain. https://dial.uclouvain.be/pr/boreal/object/boreal:229877.

Gustafsson, E. 2014. “Variation of English Passives Used by Swedes A Corpus-Based Study of the Usage of *Be*-Passives and *Get*-Passives.” Linnæus unviersity, 1-33.

Hansen, A. 1967. *Moderne Dansk*. Vol.3. Grafisk Forlag.

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), 237–258. https://doi.org/10.1111/j.1473-4192.1994.tb00065.x

Dirdal, H, Eli-Marie Danbolt Drange2, Anne-Line Graedler3, Tale M. Guldal2, Ingrid Kristine Hasund2, Susan Lee Nacey3, Sylvi Rørvik3 University of Oslo1, University of Agder2, Inland Norway University of Applied Sciences3. https://susannacey.hihm.no/wp-content/uploads/2017/09/Abstract-LCR-4-Poster-Dirdal-et-al..pdf.

Hinkel, El. 2004. “Tense, Aspect and the Passive Voice in L1 and L2 Academic Texts.” *Language Teaching Research* 8 (1): 5–29. https://doi.org/10.1191/1362168804lr132oa.

Hovdhaugen, E. 1977. Om og omkring passiv i norsk. In Thorstein Fretheim (ed.), *Sentrale problemer i norsk syntaks*, 15–46. Oslo: Universitetsforlaget. https://doi.org/10.4324/9780203133590.

Huddleston, R. D. 1976. *An introduction to English transformational syntax*. London: Longman,

Huddleston, R. D., and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language Leatherbound Edition*. Cambridge, England: Cambridge University Press.

Huddleston, R., and Geqffrry P. 2005. “The Cambridge Grammar of the English Language.” *Zeitschrift Für Anglistik Und Amerikanistik* 53 (2): 193– 94.

Huddleston, Rodney D., and Geoffrey K. Pullum. 2006. *A Student’s Introduction to English Grammar*. Cambridge, England: Cambridge University Press.

Jespersen, O. 1933. *Essentials of English Grammar: 25Th Impression, 1987*. London, England: Routledge. https://doi.org/10.4324/9780203425787.

Johansson, S. 2008. “Contrastive Analysis and Learner Language: A Corpus-Based Approach.” Norway : University of Oslo.

Lakoff, R. 1971. *Passive Resistance*. Papers from CLS 7.49-61.

Lasnik, H. & Fiengo, R. 1974. *Complement Object Deletion*. Linguistic Inquiry 5(4), 535571.

Leech, G. N., and Jan S. 1994. *A Communicative Grammar of English*. 2nd ed. London, England: Longman.

Leech, G., Marianne H., Christian M., and Nicholas S. 2009. *Studies in English Language: Change in Contemporary English: A Grammatical Study*. Cambridge, England: Cambridge University Press.

Lorenz, Gunter R. 1999. *Adjective intensification - learners versus native speakers: A corpus study of argumentative writing*. Amsterdam: Rodop.

Marchman, V. A., Elizabeth B., Antionette B., and Amy B. Good. 1991. “Functional Constraints of the Acquisition of the Passive: Toward a Model of the Competence to Perform.” *First Language* 11 (31): 65–92. https://doi.org/10.1177/014272379101103104.

McEnery, T., & Hardie, A. (2012). *Cambridge textbooks in linguistics: Corpus linguistics: Method, theory and practice: Method, theory and practice*. Cambridge University Press. https://doi.org/10.1017/cbo9780511981395.

Moreb, B. 2016. “The Frequency of the Passive Voice in Freshman Academic Books.” Nida, Eugene A. 1964. *Towards a Science of Translating*. Leiden : E.J. Brill. Orlando, Florida: University of Central Florida. 1-134

Ouhalla, J. 1999. *Transformational Grammar: From Principles and Parameters to Minimalism*. 2nd ed. London: Arnold.

Myers, L. M. 1966. The Roots of Modern English. Boston: Little, Brown & Co.

Palmer, F. R. 1965. *Linguistic Study of the English Verb*. London, England: Prentice Hall Press.

Persson, G. 1990. *Meanings, Models and Metaphors: Study in Lexical Semantics in English*. Stockholm, Sweden: Almqvist & Wiksell Internat.

Pinker, J. S. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: Mit Press.

Pullum, G. K. 2014. "Fear and Loathing of English Passive." *Language and Communication* 37: 60–74.

Quirk, R., et al. (1985) *A Comprehensive Grammar of the English Language*. London: Longman

Radford, A. 1997. *Syntax: A Minimalist Approach*. Cambridge: Cambridge University

Raimes, A. 1998. *How English Works: A Grammar Handbook with Readings*. Cambridge University Press.

Sasaki, K. 1999. The Semantics of Get-Passives. Journal of the Faculty of International Studies, Utsunomiya.

Schwarz, S. 2015. "Passive voice in American soap opera dialogue." *Studia neophilologica* 87 (2): 152–70. https://doi.org/10.1080/00393274.2015.1049831.

Shibatani. 1988. "Introduction." edited by M. Shibatani, 1–8. Amsterdam: John Benjamins.

Shintani, M. (1979). The Frequency and Usage of the English Passive. Unpublished Ph.D. dissertation, UCLA

ŠIMÍK, WIERZBA & KAMALI. 2014. Givenness and the position of the direct object in the Czech clause. *Formal Approaches to Slavic Linguistics (FASL),* 22**,** 302-318.

Stanley, J. 1975. Passive motivation. Foundations of Language, 13, 25–39

Sussex, R. 1982. A note on the Get-passive Construction. Australian Journal of Linguistics 2:8395

Svartvik, J. 1966. *On Voice in the English Verb. The Hague: Mouton*.

Thornbury, S. 2006. "An A to Z of ELT: A Dictionary of Terms and Concepts Used in English Language Teaching." *Macmillian*. University of Oslo.

Thompson, D., S. P. Ling, Andriy M., Fernanda F., and Christoph S. 2013. "Patient-Related Constraints on Get- and BePassive Uses in English: Evidence from Paraphrasing." *Frontiers in Psychology* 4. https://doi.org/10.3389/fpsyg.2013.00848.

Yannuar, N., Shitadevi, I. A., Basthomi, Y., & Widiati, U. 2014. Active and passive voice constructions by Indonesian student writers. Theory and Practice in Language Studies, 4(7), 1400-1402.