

**SUBJECT-VERB AGREEMENT OF FINITE VERBS IN THE
PRESENT TENSE IN WRITTEN PRODUCTIONS OF FINNISH
SECONDARY SCHOOL LEARNERS OF ENGLISH AS A
SECOND LANGUAGE**

A Pro Gradu Thesis

Nina Vaurula

Jyväskylän yliopisto
Kielten laitos
Englanti
Helmikuu 2012

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistinen tiedekunta	Laitos – Department Kielten laitos
Tekijä – Author Nina Vaurula	
Työn nimi – Title Subject-verb agreement of finite verbs in the present tense in written productions of Finnish secondary school learners of English as second language	
Oppiaine – Subject Englanti	Työn laji – Level Pro Gradu -tutkielma
Aika – Month and year Helmikuu 2012	Sivumäärä – Number of pages 73
<p>Tiivistelmä – Abstract</p> <p>Subjektin ja verbin välinen kongruenssi on englannin kieliopin perusasioita, joka toimii tiettyjen säännönalaisuuksien mukaisesti. Ilmiö on kuitenkin kiinnostava, sillä jopa edistyneetkin oppijat tekevät kongruenssin tuottamisessa virheitä. Aiemmat tutkimukset ovat osoittaneet, että muun muassa subjektin kompleksisuudella ja semanttisilla piirteillä on vaikutusta kongruenssin hallintaan. Aiemmat tutkimukset eivät kuitenkaan ole ottaneet huomioon oppijoiden taitotasoa ja mahdollisia eroja kongruenssin tuottamisessa eri taitotasojen välillä. Tämän tutkimuksen tarkoituksena on selvittää, miten yläkouluikäisten taitotasoeroilla on vaikutusta kongruenssin tuottamiseen englannin kielessä. Tutkimuksessa Lisäksi haluttiin selvittää, vaikuttaako subjektin kompleksisuus kongruenssin tuottamiseen oikein ja onko taitotasojen välillä eroja eri subjektien käytössä.</p> <p>Tutkimuksen aineisto koostuu 141 yläkoululaisten kirjoittamasta mielipidekirjoituksesta, jotka ovat osa CEFLING-projektia varten kerätystä aineistosta. Aineisto on jaettu taitotasoinen Eurooppalaisen Viitekehityksen taitotasojen mukaisesti. Tutkielmassa käytetyssä aineistossa tekstit ovat jakautuneet taitotasolle A1-B2. Kongruenssitapaukset etsittiin teksteistä taitotasoinen, jonka jälkeen kongruenssin frekvenssi laskettiin jokaiselle taitotasolle. Tämän jälkeen laskettiin, kuinka suuri osa tapauksista on tuotettu oikein kullakin taitotasolla. Lisäksi tarkasteltiin, miten subjektin kompleksisuus vaikuttaa kongruenssin tuottamiseen kieliopillisesti oikein ja tuloksia vertailtiin eri taitotasojen välillä.</p> <p>Tuloksista selvisi yllättäen, että subjektin ja verbin välistä kongruenssia vaativien finiittiverbien käyttö laski taitojen noustessa korkeammalle tasolle, mikä saattoi johtua siitä, että oppilaat tuottivat monimutkaisempia lauserakenteita kielitaidon karttuessa. Kieliopillinen oikeakielisyys kuitenkin kasvoi mitä korkeammasta taitotasosta oli kyse. Kompleksiset subjektit tuottivat enemmän virheellisiä tapauksia. Yksikölliset subjektit tuottivat enemmän oikeakielisiä tapauksia kuin monikolliset johtuen siitä, että yksikön kolmannen persoonan -s:ää käytettiin liikaa – myös monikollisten subjektien kanssa. Kongruenssin hallinta oli selkeästi vielä kesken, sillä saman virkkeen sisällä oppilaat saattoivat tuottaa taivutetun verbin oikein yhdessä tapauksessa ja heti seuraavassa tuottivat sen väärin.</p> <p>Tutkielma antoi viitteitä siitä, millaisia eroja kongruenssin hallinnassa eri taitotasojen välillä on. Tarkempien tulosten saamiseksi tulisi seurata yksilöllistä kielen kehitystä pikittäistutkimuksen avulla, jotta varsinaista kielen kehitystä pystyttäisiin seuraamaan paremmin.</p>	
Asiasanat – Keywords learner language, subject-verb agreement; complexity, accuracy and fluency	
Säilytyspaikka – Depository Kielten laitos	

TABLE OF CONTENTS

1 INTRODUCTION	6
2 BACKGROUND TO THE STUDY.....	10
2.1 Learner language.....	10
2.1.1 Defining learner language	11
2.1.2 Variation in learner language.....	12
2.1.3 Importance of learner language.....	14
2.2 Research on learner language.....	15
2.2.1 History of SLA research	16
2.2.2 Research methods used to analyze learner language.....	19
2.3 Complexity, accuracy and fluency in learner language	23
2.3.1 Defining complexity, accuracy, and fluency	24
2.3.2 Measuring complexity, accuracy, and fluency	26
3 PREVIOUS STUDIES	29
4 THE PRESENT STUDY	42
4.1 Aims of the study	42
4.2 Data collection	44
4.3 Analysis of the data.....	45
5 FINDINGS.....	48
5.1 Frequency of subject-verb agreement in the data.....	48
5.2 Accuracy in the use of subject-verb agreement in the data	50
5.3 Distribution of different types of noun phrases in the data.....	51
6 DISCUSSION	60
7 CONCLUSION	67
BIBLIOGRAPHY	71

1 INTRODUCTION

Even though *subject-verb agreement* is one of the basic features of the English grammar and has been referred to as a fairly simple or easy semantic feature (see Bock and Miller 1991, Eubank 1994), even advanced learners and sometimes also native speakers fail to use the feature correctly. This is especially the case with the 3rd person singular *-s* that is acquired fairly late among the different morphemes studied in English as a second language (L2). Grammatical morphemes of English emerge in *learner language* (LL) at different times, and the rates at which they are mastered vary greatly (Hsieh et al 1999: 532).

When it comes to subject-verb agreement, the plural noun *-s* is acquired much earlier than the 3rd person singular *-s* and, consequently, there have been various attempts at trying to explain the phenomenon. Eubank (1994: 84), for example, claims that the unstressed, word-final position of the 3rd person singular *-s* makes it difficult for learners to pay attention to the feature, and also the infrequent inflection of the feature makes it harder for L2 learners to acquire correct inflection. Hsieh et al (1999: 532), for their part, suggest that in English the plural noun *-s* is semantically more important than the inflection of verbs and that in the earliest stages of grammatical development lexical categories outweigh grammatical categories. Therefore learning plural nouns have more meaning to learners than learning to inflect verbs as the plurality conveys more meaning than adding a 3rd person singular *-s* to the verb. Some other studies, such as Bock and Miller (1991) have looked at the features of the noun phrases (NP) to see how they affect the correct use of subject-verb agreement, whether it might be semantic or syntactic features of NPs.

However, previous studies have concentrated on subject-verb agreement only from a linguistic point of view and have completely ignored learners' language skills. As Ellis and Barkhuizen (2005: 79) suggested, learners should be grouped

according to their *proficiency levels* in order to measure language development at the same stage of proficiency. Therefore the present study both examines subject-verb agreement in L2 learners' language use and takes into account the learners' language proficiency. The aim is to look at how frequently and accurately Finnish L2 learners of English use subject-verb agreement in their written productions and compare the results between different proficiency levels. Another point of interest will be to analyse different types of NPs to see whether their complexity has an effect on how accurately the learners use subject-verb agreement. A comparison of the different NPs is also done between different levels of language proficiency.

The present study is a part of a wider Cefling project (Linguistic basis of the Common European Framework for L2 English and L2 Finnish) funded by the Academy of Finland, which aims at combining research in second language acquisition (SLA) and in language assessment (<http://www.jyu.fi/hum/laitokset/kielet/cefling/en>, 4.2.2012). The project is a part of one of the most significant European research network *Second Language Acquisition and Testing in Europe* (SLATE) in its field. The main aim of the Cefling project is to find out how learners' language skills develop from one stage to another and it is based on two assumptions: 1) language skills develop through stages and these stages can be separated and described, and 2) similar processes and features between different learners' language skills can be found at a certain stage. These stages and their features have been introduced to language education through descriptive scales as a part of *the Common European Framework of References for Languages* (CEFR) (http://www.coe.int/t/dg4/linguistic/cadre_en.asp, 4.2.2012) and they have become the yardsticks of language curricula, examinations, materials and courses in Europe. Language development is divided into six different scales ranging from A1 to C2 (for more information, see the web pages above). The present study is linked to the Cefling project as it uses the same data and

focuses on comparing language proficiency at different scales defined in the CEFR.

The methodological framework of the study relies on the DEMfad model (Franceschina et al. 2006) that has also been used in the Cefling project. With the model, different dimensions of language development are examined and therefore it enables reliable comparison between different levels, domains and languages. The model focuses on observing the features of fluency, accuracy and complexity from the learners' written productions. SLA studies have basically been divided into three different categories: descriptive studies, experimental pedagogical studies, and hypothesis-testing studies (Lightbown 1985, as quoted in McLaughlin 1987: 2). Descriptive studies rely on collected samples of LL from L2 learners and speakers which are then compared to the target language (TL), and thus the purpose is to find similarities and differences in language use between L2 and TL. Experimental pedagogical studies, for their part, try to manipulate certain variables, such as instructions or learning conditions, experimentally to define their effects on classroom learning. In the hypothesis-testing studies a hypothesis is always stated in the beginning of the study and it is based on the findings of previous studies. The hypothesis is then tested in the study either to confirm or prove it wrong. Fundamentally, all research is hypothesis testing as all studies base on theories, and without a theory, there cannot be hypotheses. The present study is a descriptive study as it relies on written samples collected from Finnish L2 learners of English.

The present study sets of by presenting the background to the study in Chapter 2 by introducing some of the most significant methods used to study SLA as well as explaining the main terms used in the present study. Chapter 3 moves on with an introduction of some of the main studies prior to the present study and also summarises their results. The aims of the present study are demonstrated in Chapter 4 with more detailed information on the data and

ways of analysing them. The findings of the present study are reported in Chapter 5 which are then interpreted and discussed in Chapter 6. In the final chapter the key findings of the present study are summarised, and the present study is also evaluated with suggestions for further research on the topic.

2 BACKGROUND TO THE STUDY

Although the history of SLA research is fairly young, numerous studies have been made on how languages are learned, and many studies have also focused on learning certain features of TL. Different schools have their own ideas of how the process actually happens, and also different methods to study the emerging language have been used. Studying LL itself has been seen important in not only understanding the development but more importantly the processes and stages behind the development. As a matter of fact, according to Ellis and Barkhuizen (2005: 5), the goal of SLA research for many researchers is to describe the language development of L2 learners and to explain how it develops over time.

The following sections will give an introduction on what is meant by LL and what have been the key points in researching it. In section 2.1, LL is defined in greater detail, and the importance of studying and understanding it is explained as well. Section 2.2 presents the background of LL research. It introduces briefly the history of SLA research and also the research methods used to study LL. Section 2.3 creates the methodological basis for the present study as it introduces the measures used in the method of analysis in this study.

2.1 Learner language

LL is oral or written language produced by L2 learners, and it serves as the primary data for the study of SLA (Ellis and Barkhuizen 2005: 4). It is used to describe and explain L2 learners' language competence and how this develops over time. It also aims at describing L2 learners' language skills and it is an independent system of both their first language (L1) and L2 (VanPatten and Benati 2010: 100). In the following, LL and its variation are defined and also the importance of LL research will be explained.

2.1.1 Defining learner language

LL, also known as interlanguage (IL), is a term launched by Selinker in 1972, and with it he refers to temporary grammars constructed by L2 learners on their way to mastering TL (McLaughlin 1987: 60). It is the learners' prevailing system of TL at a single point in time and it also covers the range of interlocking TL systems that describes the learners' development over time. Ellis (1986: 42) defines LL as the systematic knowledge of languages in common that is separate from both the learners' L1 and the TL system they are trying to learn, and it follows a universal route that is influenced by the learners' age, background and environment.

LL is a production of five central cognitive processes: 1) language transfer 2) transfer of training 3) strategies of L2 learning 4) strategies of L2 communication 5) overgeneralization of TL linguistic material (Selinker 1972, as quoted in Ellis 1986: 48). It evolves over time getting closer to TL system and it can be found both in children's and adult learners' language use. The LL system is also dynamic as it is constantly changing. Tarone (as quoted in McLaughlin 1987: 64) introduced a capability continuum: LL develops away from L1 towards L2. It has different features from both languages at different points in the continuum but still the development is systematic like in any other natural language with the same universals, and it also has a syntax of its own. The rules are not fixed but they are under constant development as the knowledge of TL increases with new information about it. Errors, on the other hand, when they occur, are not mistakes but actually evidence of this rule-based behavior with linguistic rules and principles (White 2003, as quoted in Doughty and Long 2003:19).

Languages are generally systematic in the sense that there are regularities which can be identified and without which communication could not take place (Tarone 1988: 1). Researchers look at regularities in different levels, for example, in syntactic, pragmatic, and interactional levels and thus have different views

on both regularities and variation. However, systematicity is not always obvious as spoken language differs in many ways from written language which is considered to be the normal grammatical and semantic system of Standard English. LL is seen to be systematic because it has, like any other language, its own rules and it must be studied by standard methods used in linguistic enquiry (Corder 1971, as quoted in Tarone 1988: 8). On the other hand, it is unstable and creative with its very unique rules.

However, many learners fail to reach TL competence because of language fossilization. According to VanPatten and Benati (2010: 100), approximately 95 per cent of all L2 learners will never reach native-like competence in TL. L1 development differs from L2 development as it cannot be fossilized whereas fossilization in L2 can be caused by L1 transfer (Selinker 1972, as quoted in Ellis 1986: 61). Moreover, communication strategies in L2 may cause the learner to stop learning the language when they have reached a point where they can communicate in L2 – it is not as necessary to learn anymore.

2.1.2 Variation in learner language

As discussed in the previous section, it is very typical of LL to be changing fast as learners do not hear a variable like their own and do not thus get reinforcement to the forms and structures they produce (Larsen-Freeman and Long 1991: 82). Rather they try to change their language to resemble more the variant they hear. So instead of discussing the differences between learners' L1 and L2, the focus should be on the variability of LL. This variation has been defined 1) as internal variation in LL independent of the TL system, and 2) as a kind of external variation that is determined by the use/non-use of predetermined TL forms in obligatory contexts (Tarone 1988: 11).

There are two basic types of LL variation: *systematic and non-systematic* (Ellis 1985b as quoted in Tarone 1988: 19). *Systematic variation* is variation that can be explained and predicted. One type of systematic variation is individual

variation (internal variation) which involves different performance of different individuals in their production of IL, both at a single point in time, and over time. This is caused by individual learner factors such as attitudes and motivation. The other type of systematic variation is contextual variation (external variation) where situational contexts determine contextual variation which can be related to the nature of the task, topic, interlocutor, physical location, and so on. Contextual variation can also be determined by the linguistic context, including phonetic variation. By contrast, *non-systematic or free variation* is random and unpredictable, and utterances that are produced within minutes of each other in the same situational and linguistic context might have no apparent pattern in their use.

Another possible cause for variation in LL is that language is used for various purposes and to serve different functions, and therefore also the forms used to perform those different functions will also often vary as a consequence (Tarone 1988: 94). LL may often have rules that differ from those of TL and thus LL forms may appear to differ from those of TL rules and norms. After all this might in fact be quite invariably related to language function – good examples of this are memorized chunks or long lexical items which are used to express the appropriate functions long before they even become analyzed into their components.

The context can as well have an influence on L2 learners' language production. When given the time to pre-plan their production, L2 learners will produce more complex grammatical constructions than, for example, in a spontaneous situation (Ellis and Barkhuizen 2005: 22). They might also switch from L2 to L1 or even mix the languages, and their mastery of certain grammatical structures can vary depending on the type of task used to elicit TL forms.

LL can also vary from task to task (see Larsen-Freeman 1975a, b and Lococo 1976), however, the causes for the variability are not clear. Task effect on learner accuracy has been found in many studies which have not even been designed to

study variation in LL and thus the variability has been unexpected (Tarone 1988: 14). The nature of the task causes unpredictable variation in the learners' productions and both the LL system as it appears in the task and the variation caused by the change in tasks should be analyzed, and comparison of similar tasks between different developmental stages should be made (Trévisé and Porquier 1986, as quoted in Tarone 1988: 18). Different tasks have different demands on the learner as they are designed differently – learners need different skills in different types of tasks (Larsen-Freeman and Long 1991: 31). Some tasks require more communicative skills and some need more focus on form. When the focus is on form, the language is usually more careful and thus TL forms are more varied (Tarone 1983, as quoted in Larsen-Freeman and Long 1991: 32). On the other hand, certain types of tasks can induce certain types of mistakes because of a different focus. For example, in translation tasks, where the focus is more on content, L1 influence can be more effective. Therefore there is a need for cross-research between similar tasks (Larsen-Freeman and Long 1991: 33).

2.1.3 Importance of learner language

In the 1940s, 1950s and 1960s LL was thought to be influenced by the learner's L1 and it was therefore believed that many errors were caused by falling back to the syntax of their L1. According to Lado (1957, as quoted in Tarone 1988: 8), learner errors could actually be predicted if both L1 and TL systems were analyzed carefully and compared for differences. Contrastive analysis (CA) also assumes that errors occur due to differences in L1 and TL.

Many researchers in the field of SLA are particularly interested in how learners learn or acquire a new L2 and how they use the language (Ellis and Barkhuizen 2005: 15). LL functions here as a source of information as TL learners are the best experts of their own minds. One way of researching LL is to ask learners to report on their own learning. They can be asked to describe the linguistic system that they base their language use on or to explain the processes and

factors involved in acquiring TL. Another way is to collect samples of LL – what learners know about the language is best reflected in their comprehension of input and in the language they produce. The samples should be collected in a context that resembles that of a natural context, i.e. in natural language activities. The learners can then be asked to comment on their own material and to ponder why they made mistakes and what they did differently when they succeeded. The latter question is especially problematic as learners sometimes tend to make errors more than at another time as they might first use the form correctly but only a little later they make an error with the exactly same matter.

LL research has basically two goals: 1) it tries to describe the linguistic systems of the learner at different stages of development, and 2) it attempts to explain the processes and factors involved in acquiring L2 (Ellis and Barkhuizen 2005: 15). Actually, one of the main goals of SLA research is to reveal learners' implicit language knowledge – the knowledge that is unconscious and available for automatic use. This would truly shed light on the different stages of language acquisition and how languages actually are acquired. Moreover, it aims at providing evidence for developing and testing LL through different theories of SLA. Learner errors are also a significant source of information as they function as research evidence about how languages are learned, and, more importantly, they show teachers what and what not their learners master. The following sections will focus on LL research as it discusses briefly the history of LL research as well as introduces the key methods used to study it.

2.2 Research on learner language

SLA is a study of how languages are learned and it has a variety of starting points ranging from linguistics to sociology and psychology (Ellis and Barkhuizen 2005: 3). It refers to the acquisition of any language that is acquired after acquiring L1 irrespective of it actually being the second, third or even 4th language. Acquiring L2 is different when compared to acquiring L1 as all

learners have already once acquired a language and thus have an idea of how languages work (Lightbown and Spada 1993: 21). On the other hand, the knowledge of other languages can also interfere with L2 thus resulting in incorrect utterances and errors. SLA is, however, a relatively young study of field as research papers concerning its study have started to appear only in the 1960s. The history of its research is summarized in the following section followed by an introduction of the methods used to study LL.

2.2.1 History of SLA research

SLA research has its roots in two seminal publications by Corder (1967) and Selinker (1972). Corder was concerned with language instruction, and introduced the idea of an 'internal syllabus' which meant that all L2 learners would have a similar kind of internal system for language acquisition as what children acquiring their L1 do (as quoted in VanPatten and Benati 2010: 2). This syllabus might differ greatly from the instruction learners received and therefore he also considered it important to distinguish input from intake. Selinker, for his part, introduced the term 'interlanguage' indicating that learners have an internal linguistic system that is neither L1 nor L2 but something in-between. He also introduced two central terms 'L1 transfer' and 'fossilization' that are still used today in SLA studies.

In the 1970s descriptive studies dominated the field attempting to challenge behaviourism and seeding the ideas of Corder and Selinker (VanPatten and Benati 2010: 3). *Error analysis* (EA) that examined learners' output emerged in the field accompanied by the idea of an acquisition order (or morpheme studies). A pioneer of the morpheme studies was Brown with his study (1973) on children acquiring their L1 (as quoted in McLaughlin 1987: 66). He found out that children had a certain order in which they acquired the different morphemes, and most of the later SLA studies on morpheme acquisition have referred to Brown's pioneering study. On the other hand, attention was focused on the role of input and especially on communicative input (VanPatten

1990, as quoted in VanPatten and James 1990: 19). It was not, however, clear whether input actually was enough to result in changes in a learner's linguistic system, or whether it should be simplified or comprehensible. Language was also seen to develop through transitional stages which pointed to the importance on L1 studies (VanPatten and Benati 2010: 27). The stages do not, however, follow each other systematically but rather learners move from one stage to another which also causes overlapping.

The early 1980s were led by Krashen's and Chomsky's ideas of language acquisition. Krashen introduced the Monitor model, input hypothesis and 'acquisition' versus 'learning' that ended up receiving much criticism as they left many questions unanswered (VanPatten and Benati 2010: 3). The basic idea in Krashen's theory is that learning becomes acquisition but it cannot be tested empirically, and all in all, the empirical basis of his theory is weak. Chomsky introduced a Language Acquisition Device (LAD) that, according to him, describes a child's initial state before any linguistic information is received from the environment (as quoted in McLaughlin 1987: 23). The LAD contains linguistic universals with which grammar is generated from the received input and thus the rules of a language are acquired. Krashen claimed also adults to have the LAD but adults are not at the initial point anymore having already received years of linguistic input and therefore no longer have access to the device. Therefore some other mental process needs to be responsible for later language acquisition, and as language was also considered to be uniquely human, the Universal Grammar theory assumes that an unconscious universal ability to 'cognize' language, as Chomsky puts it, underlies all language acquisition.

SLA was also seen as a complex cognitive skill as acquiring a language requires much rehearsing and repeating (McLaughlin 1987: 134). Therefore the psychologists and psycholinguists behind the Cognitive theory suggested that acquisition is cognitive and involves internal representations that regulate and guide our performances. Language wise this would mean selecting appropriate

vocabulary and being able to use the language grammatically and pragmatically correctly. This requires, however, a lot of capacity and thus part of the language use needs to be automated for processing new information: automatic processing releases capacity for processing new and/or more complex linguistic information as the automated parts are retrieved to use from the memory easily and rapidly. Nevertheless, there are some problems with the validity of the theory. Firstly, it does not have a very clear theoretical ground and it has quite a specific focus when compared to other schools of language learning and secondly, and more importantly, it does not explain the acquisition of a language by itself – it is yet another way of looking at language acquisition.

In the 1990s two major approaches dominated the field. Firstly, the linguistic theory held on to the IL theory and was concerned with learners' internal mental representations and what constrained it (VanPatten and Benati 2010: 4). Secondly, the psychological camp approached SLA through what learners did with language rather than what their underlying knowledge was, and hence language was seen as normal human behaviour with nothing special in it. This became known as the Sociocultural Theory and it saw linguistic and cognitive theories being too much oriented on the mind and instead saw the learner as an active learner in a social context. Even today linguistic and cognitive approaches continue to dominate the field and they have been applied to a range of studies from L1 acquisition to natural language processing and therefore the focus of SLA by large still concentrates on the mind.

As the spectrum of language acquisition theories is wide and as they all have their limitations, it cannot be said that there would be one theory to explain it all. Different schools have their own ways of interpreting language acquisition and new ways of understanding the phenomenon emerge now and then. Language acquisition is after all such a complex process that truly understanding it requires much research and testing. To be able to grasp all the

processes behind language acquisition would certainly change the world of learning and teaching languages.

2.2.2 Research methods used to analyze learner language

In analysing LL three methods have dominated the field: *error analysis*, *obligatory context analysis* and *target-like use*. They all include performance analysis of samples of LL and, for their own part, have brought new evidence of how languages are acquired. However, none of the methods have been without problems or criticism. These three methods will be presented in more detail in this section.

Error analysis (EA) is a study of errors in learners' speaking and writing (Ellis and Barkhuizen 2005: 51). It was preceded by CA that sought to predict the errors that learners make (Ellis 1994: 47). The errors were predicted by comparing the learners' L1 to TL and the underlying assumption was that L1 will interfere with L2 when the patterns in L1 differ from those in L2. When this assumption became challenged, EA then provided a method for investigating LL per se: in EA the samples of LL form the data for the analysis from which errors are being identified, described, explained, and evaluated. By analyzing errors it could be seen how much learners had acquired and this also helped them to discover the rules of the language (Corder 1967, as quoted in Ellis 1994: 48). As a result, EA brought evidence to researchers of how languages were acquired. However, a variety of factors can influence the errors that learners make (Ellis 1994: 49). For example, different L1s can result in different errors as L1 systems are different from each other, and also different conditions, such as the medium of production and the topic under which language is produced, might affect the type of errors learners make (see for example Lococo 1976).

EA does not however provide a complete picture of how L2 is acquired as it describes only the errors that learners have made. The heaviest criticism has been laid on the weaknesses in its methodological procedures and its

limitations in scope as it fails to provide a complete picture of language learning: it does not take into account stages of development and as a result, although categorising errors may have brought important information to language teaching, it has not been so successful in helping to understand how learners develop their knowledge of L2 over time (Ellis 1994: 68). Moreover, the researchers need to decide in each case, what constitutes as an error before they are able to recognize them, how to describe and categorize them, how to explain the reasons behind them, and also which variety of TL functions as the norm. This can cause fundamental differences in results from one study to another. According to McLaughlin (1987: 68), another problem is that the studies have in most cases been cross-sectional, which does not actually tell whether the errors are prevalent at specific points in language development or whether certain errors persist longer than others.

The second dominating method in the field is the *obligatory context analysis* and it has been widely used in L2 studies. The idea behind it is that grammatical morphemes are obligatory in certain contexts and thus by calculating the correct use of the morphemes, it comprises a method for examining how accurately learners use these linguistic features (Ellis and Barkhuizen 2005: 73). A comparison is made between the forms used by L2 learners and TL norms. L2 learners create obligatory contexts for certain linguistic features in their TL production but they may not always supply the feature (Ellis 1994: 74). According to Brown (1973: 296), each obligatory context is a test item that is passed by supplying a correct morpheme, and failed by supplying an incorrect one or no morpheme at all. A criterion for acquisition is therefore considered to be output-where-required at a level of 80-90% correct suppliance. This has been considered to function as a 'norm' in many studies as they refer to Brown's study as the guideline for later studies. The analysis begins by a collection of samples of LL (Ellis 1994: 74). From the samples obligatory occasions are then identified and the percentage of accurate use is calculated. Accuracy here

means whether a morpheme has been supplied on all occasions where required. The calculation of accuracy for individual morphemes is presented in Figure 1.

- 1) Determine which morpheme is to be investigated
- 2) Go through the data and identify obligatory occasions for the use of the morpheme.
Count the total number of occasions
- 3) Establish whether the correct morpheme is supplied in each obligatory context.
Count the number of times it is supplied
4. Calculate the percentage of accurate use with this formula

$$\frac{\text{n correct suppliance in contexts}}{\text{total obligatory contexts}} \times 100 = \text{per cent accuracy}$$
5. Repeat the procedure for the other morphemes to be investigated

Figure 1. Calculating accuracy for individual morphemes (Ellis and Barkhuizen 2005: 80)

The analysis is, however, not without problems and has received criticism. Long and Sato (1983, as quoted in Ellis and Barkhuizen 2005: 79) claim that the analysis does not tell anything about if the learners understand the functions of the morphemes they have acquired. Moreover, the analysis model does not take into account overgeneralization, meaning that it leaves out the occasions where a feature has been supplied in a context where it is not obligatory (Ellis 1994: 75). Acquisition should therefore be measured as knowing when and when not to use a certain linguistic feature. Analysts should also be careful to be consistent in their decision making and make their coding decisions explicit. Each decision should also be provided with a rationale. Ellis and Barkhuizen (2005: 79) suggest a few ways of how to come around some of the problems of obligatory context analysis. Firstly, the set of morphemes should be expanded as it is usually around 12 to 14 morphemes and the morphemes should also be categorized. Secondly, learners should be grouped according to their proficiency so that the accuracy of morphemes could be examined with learners at the same stage of development. Moreover, a method of analysis that would also take into account both correct suppliance and overuse of a morpheme should be adopted in the research.

The problems and criticism confronted by obligatory context analysis were solved by Pica in her piloting study (1983) for a new scoring method in which

the incorrect suppliance was added to the correct use and the percentage of the correct use then calculated (VanPatten and Benati 2010: 158). The pattern for the model can be found in Figure 2.

$$\frac{\text{n correct suppliance in context}}{\text{n obligatory contexts} \times \text{n suppliance in non - obligatory contexts}} \times 100 = \text{per cent accuracy}$$

Figure 2. Calculating accuracy for target-like use analysis (Pica 1983:81).

This third method, *target-like use*, is concerned with how well learners can produce particular linguistic features. It also takes into account overuse of the linguistic feature and thus gives more reliable information on the level of mastery of the feature (Ellis and Barkhuizen 2005: 80). The method also tries to answer how, for example, the linguistic nature of certain morphemes affects the way in which they are used in obligatory and non-obligatory contexts, and how the context influences their use. Moreover, it aims at shedding light on how instruction contributes to the acquisition of morphemes.

This new method was considered to be a major improvement in scoring accuracy (VanPatten and Benati 2010: 158). However, accurate use of a certain feature of TL does not necessarily mean that the learner knows the function of the form but has just learned to use it as a chunk (Long and Sato 1984, as quoted by Goldschneider and DeKeyser 2001: 38). According to Dulay and Burt (1980, as quoted by Ellis and Barkhuizen 2005: 80), any attempt to supply a morpheme, even if it was incorrect, should be taken as some evidence of a greater degree of acquisition than failing to supply any. A qualitative study of language performance would therefore be more insightful than a quantitative one that just observes the numbers and ranks learners accordingly.

Both obligatory context and target-like use analysis compare LL to TL.

Therefore there is the danger of 'comparative fallacy' which means that these analyses ignore the fact that L2 learners' language is a unique system in the process of just learning TL (Bley-Vroman 1983, as quoted in Ellis 1994: 75).

Learners might not be showing the whole range of their language competence at the measured point and their performance can change depending on the context and situation (Ellis and Barkhuizen 2005: 22). Sometimes they make errors using a particular TL form and sometimes they use the form correctly. One way to overcome this difficulty could be to study the developmental patterns that learners have at different stages of LL and to calculate the frequency with which they use these patterns at those stages (Ellis 1994: 75). A longitudinal study could therefore give more reliable results on language development at different stages.

In this chapter the history of LL research was summarized and also the main trends in LL studies were put together. As was noticed, different methods of studying SLA are used widely but they are not, however, without problems. Not one method to reveal all the secrets behind SLA has yet been discovered. The history of SLA research is rather short, and new, better methods are being sought all the time. The focus will now turn to measures of complexity, accuracy and fluency that are used to study L2 learners' general language performance. They are also used as measures in the present study.

2.3 Complexity, accuracy and fluency in learner language

Complexity, accuracy and fluency (CAF) have been used as measures of language proficiency in many studies of SLA and they are used for describing language performance, for example, describing proficiency levels and task features (Pallotti 2009: 590). They have also been used to assess oral and written performance, and work as indicators of learners' underlying language proficiency as well as measures in language learning (Housen and Kuiken 2009: 461). The origins of the triad lie in the L2 pedagogy of the 1980s when a difference between fluent and accurate L2 usage was made. Complexity was added in the 1990s by Skehan (see Skehan 1998). This chapter focuses on defining CAF and it also explains how they are measured in LL.

2.3.1 Defining complexity, accuracy, and fluency

CAF are classified as separate aspects of development and this classification includes three types of intuitions of especially L2 learners' written language as they become more proficient: they write more fluently, or write more in the same amount of time; they write more accurately, or produce fewer errors in their writing; and they write more grammatically and lexically complex sentences (Wolfe-Quintero et al. 1998: 4). However, individual variability (see section 2.1.2) should be taken into consideration when assessing the results. There can, for example, be learners who are fluent but inaccurate or those who are accurate but non-fluent. On the other hand, an increase in one may occur at the expense of the others. In a more traditional sense, complexity refers to using a wide range of structures and vocabulary, accuracy to error-free production, and fluency to native-like rapidity (Lennon 1990, as quoted by Wolfe-Quintero et al. 1998: 4). These three aspects will next be explained in more detail.

The first part of the triad, *complexity*, and more precisely syntactic complexity, refers to the L2 learner's syntactic repertoire and the ability to utilise this in a variety of situations (Ortega 2003: 492). The length of production units, the range of structural types and the amount of embedding have also been considered to be important measures in defining the levels of complexity. These measures have been used to assess task-related variation in L2 writing and the differences between learners on different proficiency levels. It should, however, be noted that more complex does not necessarily mean better or more developed, but the learner's development of discourse and sociolinguistic repertoire should be on a level that enables the learner to adapt language appropriately in different situations. Complexity is anyhow the most problematic of the CAF triad as it has various definitions because it is used to refer to both task properties and language performance, and even though it was restricted only on assessing language performance, it would still have multiple meanings in different aspects of language and communication: the same structure can have, for example, lexical, interactional, propositional, and

various types of complex grammatical aspects and is thus more difficult to comprehend (Pallotti 2009: 592). Also the concept of 'complex' is multifaceted: it can be purely structural (composed of two or more parts), it can refer to multiple lexical or syntactic alternatives of language use (cognitively demanding), or it can be defined as the ability to use more advanced language (appearing later).

Secondly, *accuracy* refers to the degree of deviation from TL norms which are usually characterized as errors (Housen and Kuiken 2009: 463). Therefore accuracy itself does not indicate language development per se but it compares language productions with target-like use, i.e. the learner productions, and the errors in them (see section 2.2.2), to the norms of TL (Pallotti 2009: 592). Errors can however be weighed differently or they might be graded according to their developmental sequence but this does not result in more accurate productions but just productions of which some are more developed or advanced.

Moreover, it is not even always clear what should be considered the norm – should it be the standard norm embodied by the native speaker of TL or a non-standard or even a non-native norm that is used in certain communities or contexts (see for example Polio 1997, Ellis 2008, James 1998).

The third aspect, *fluency*, is the ease with which L2 learners can access the language items that they need depending on the context and abilities (Lennon 1990, as quoted in Wolfe-Quintero et al. 1998: 13). Fluency includes the use of routines, such as automated chunks or pragmatic formulas, and using these kinds of different routines increases the measures of fluency because language use is easier for the learner. It has also been described as being able to produce language at a normal rate without undue pausing or hesitation (see Skehan 2009, Ellis and Barkhuizen 2005).

These three aspects can also be seen corresponding to language representation and language access (Wolfe-Quintero et al. 1998: 4). Thus complexity and accuracy would reveal the L2 learner's current level of language development

(representation), but fluency on its part would control the access to that knowledge. More recent theories anyway suggest that representations and access depend on how knowledge is encoded: as connections, instances or chunks. However, according to Ellis (1996, as quoted by Wolfe-Quintero et al. 1998: 5), language representation and language access derive from the learning process which is about the gradual strengthening of memorizing language chunks. The source of all language learning is thus dependent on memorizing these chunks, and the better the memory, the better the language learning is because there will be more representations available for restructuring new language information and accessing old information automatically. Thereby fluency measures the observable outcomes of automaticity of access, and accuracy and complexity measure the observable outcomes of representations and restructuring. With limitations in the working memory learners might, however, choose to prioritise either the context or the form to get the task done, and the more complex the task, the more difficult it is to pay attention to both (Ellis and Barkhuizen 2005: 141). This might tempt them to choose more simplified language that is already more automatic and thus safer to use instead of trying to put more sophisticated language into use that they cannot yet control sufficiently, which undoubtedly results in reduced accuracy.

Complexity, accuracy and fluency can also apply to different linguistic levels, such as phonology, morphology, syntax, discourse, and pragmatics (Wolfe-Quintero et al. 1998: 5). For example, the discourse abilities of a learner can be evaluated in terms of how well the discourse fits the context (accuracy), how varied language is used (complexity), and how rapidly or fluently the varied language is used (fluency).

2.3.2 Measuring complexity, accuracy, and fluency

SLA literature basically includes two types of developmental studies: developmental sequence studies and developmental index studies which both are used when measuring the CAF triad (see for example Bardovi-Harlig and

Bofman 1989). Examples of the first are the studies of morpheme acquisition order which are based on error and performance analysis, and defining stages in language development (Wolfe-Quintero et al. 1998: 2). Examples of the latter are studies which have measured certain features in a language sample, such as the length of error-free T-units, the number of dependent clauses per total clauses, or the uses of a certain structure per word. The mean length of utterance (MLU) has also been used as this kind of an index, but no one measure has actually proved itself to be functional enough as the results have been too mixed and imprecise. These different methods of measuring CAF are presented in the following.

The original goal for the developmental index studies was to create an index that could work as an objective measure when determining more precise descriptions of L2 developmental levels (Wolfe-Quintero et al. 1998: 3). This would make comparability between different studies and TLs easier. The developmental index would also give evidence on the influence of pedagogical instructions in different contexts and how the language use of learners differ in these different contexts.

Measures of fluency include, for example, frequency count of a particular feature, structure, or unit (Wolfe-Quintero et al. 1998: 9). A frequency score has been calculated by counting the number of words, clauses, or T-units. This can however be problematic as, for example, the number of words is connected to the time allotted to the task or to the nature of the task, and thus the numbers are not comparable with other population or between different tasks. Therefore a more valid type of measurement is a ratio measure where one type of a unit is compared to the number of another type of a unit or divided by the total number of comparable units such as the T-unit analysis. The T-unit analysis has nevertheless been quite popular also in measuring syntactic complexity of both speech and written samples (Gaies 1980: 53). It has been designed to bring out developmental patterns in syntax development and has been considered to be a satisfying and stable index of language development. Some researchers have

also created their own formulas to count the scores according to their points of interests (Wolfe-Quintero et al, 1998: 10). Fluency measures also reveal the ease with which the learner can produce language, meaning the automaticity of their language use.

Having presented the main ideas behind the CAF triad that are also used as measures in the present study, the next chapter will focus on summarising some key findings on previous studies on subject-verb agreement and the different features that might influence learners' performance on supplying the feature correctly.

3 PREVIOUS STUDIES

So far the paper has presented some theoretical features behind LL by having a look at its different aspects and points of interest in SLA research. The focus will now turn to reviewing previous studies made on subject-verb agreement and the CAF triad. The latter studies have been chosen with a focus on morphological inflection such as the 3rd person singular or the number of NP. The majority of the studies on subject-verb agreement have focused on language acquisition at early and intermediate stages but more recently studies have taken a larger focus on the advanced learners as findings on SLA at higher levels of acquisition can benefit SLA professionals and instructors better (Bardovi-Harlig and Bofman 1989: 18). Therefore also the previous studies presented here focus more on the advanced learners of English as L2. L1 of the learners have also varied across the studies but researchers have noticed that it does not affect the results significantly. Moreover, the reasons behind overuse and omission of the 3rd person singular –s have interested researchers as even advanced L2 learners of English keep omitting the feature in their productions (Tarone 1985, as quoted in Tarone 1988:97). In addition, various features of NPs have been studied to find the factors affecting the occurrence of agreement errors.

The relationship between syntactic complexity and overall accuracy was studied by Bardovi-Harlig and Bofman (1989: 17-34) in written productions of 30 advanced adult L2 learners. The learners represented five native languages (Arabic, Chinese, Korean, Malay, and Spanish) and they were chosen on the basis of having a score of approximately 550 on the TOEFL test. Each language group had six learners: three who had passed a placement exam to a university and three who had not passed it. The participants had 45 minutes to write a composition on a nontechnical topic such as “The person who most influenced your life”. Written texts were used because it is possible to evaluate the writing

of advanced learners in more detail than oral productions, and it has also been found that written and oral samples are fairly comparable.

The compositions were scored along two parameters: syntactic complexity and surface errors. Complexity meant a multicausal sentence that exhibited subordination and it was calculated as the number of clauses per T-unit. Errors were graded as syntactic, morphological, or lexical-idiomatic. Syntactic errors included errors in word order, combining sentences, omitting major (subject, verb, or object) and minor (adverb placement) constituents, and sentence combining. Morphological errors were errors in grammatical morphemes such as errors in nominal morphology (plural, number agreement), verbal morphology (subject-verb agreement, passive formation), determiners and articles, and prepositions.

The findings were quite interesting. Firstly, both the pass and the non-pass groups showed similar scores in complexity when measured by the number of clauses per T-unit, but the non-pass group scored slightly higher than the pass group but the difference was not significant. As a matter of fact, the highest complexity scores were achieved by those who did not pass the exam. Secondly, the pass group made fewer errors in all the categories than the non-pass group but the errors showed the same distribution in all the error types in both groups. Errors in grammatical morphemes were the most common in both groups and errors in syntax were the least common. The native language of the learners showed no significant differences between the participants but all showed very similar error profiles. Consequently, the results showed that the participants' IL had quite a strong syntax but their acquisition of grammatical morphemes was still fairly incomplete as they comprised the largest amount of all the errors. This suggests that SLA goes through the same route regardless of L1 even though it might show unevenness in some areas, such as syntax and morphology.

The effects of the nature of the sentential subject on subject-verb agreement errors were studied by Bock and Miller (1991). They were interested to find out what agreement errors might reveal about the features of sentential subjects. They conducted three experiments in which they wanted to find out how laboratory settings would affect the occurrence of agreement errors, whether the structural features of NP would make a difference, or whether the plurality or the position of NP would induce more errors.

In their first experiment they aimed at inducing agreement errors in laboratory settings. The participants were 40 undergraduate students at the Michigan State University. The material consisted of 32 sentence beginnings. Half of the sentences had a prepositional postmodifier after the head and the other half had subject- or object-relative clause postmodifier. There were eight versions of each of the sentence beginnings with short or long postmodifiers. Also, half of the postmodifiers were created so that they created a mismatch with the head as with a singular head followed by a plural modifier, or vice versa. In addition, they had 56 filler sentence beginnings with simple NPs (half singular, half plural). Altogether the material consisted of 88-item list put together from the sentence beginnings which were then recorded for the participants. The participants heard the sentence beginnings one at a time and they were supposed to repeat the beginnings and complete them as they saw fit. The participants were given no instructions on how to complete the sentence beginnings. The responses were scored right if the participants repeated the sentence beginnings correctly, said them only once, or produced an inflected verb that correctly differentiated singular from plural forms. Agreement errors were scored if they met all other criteria except for failing to agree with the number of NP. Uninflected verb forms were scored when all other criteria was met except for not differentiating singular from plurals. Other reproduction errors were scored as miscellany.

The results showed that 57.6 per cent ($n = 737$) of all responses were correctly produced, 4.9 per cent ($n = 63$) had agreement errors, 18.8 per cent ($n = 241$) had

uninflected verbs, and 18.7 per cent ($n = 239$) were miscellaneous responses. The majority of agreement errors occurred in mismatch situations but there were no significant differences between short or long postmodifiers. However, the participants made significantly more errors when the head was singular and the postmodifier plural than the other way. Furthermore, prepositional postmodifiers resulted in more subject-verb agreement errors. Errors in the inflection of verbs were fewer when the head was plural followed by a long postmodifier regardless of match or mismatch. This could be explained partly by the fact that also past tense forms were included in the test. Their use is much easier as they are number-neutral verb forms. 74 per cent of the miscellaneous responses occurred when the head was followed by a long postmodifier which indicates that it was more difficult for the participants to repeat longer utterances than shorter ones. Yet this had no significant effects on agreement errors which would indicate that the distance between the head and the verb form has little consequences for the agreement process.

In the second experiment they wanted to find out if the concept of NP would make a difference and therefore systematically varied the concept of NP and the concreteness of the postmodifiers. In this experiment there were 64 participants from the same university and they were recruited via an advertisement and offered three dollars each for taking part in the experiment. The material consisted of 32 sentence beginnings with the head followed by a prepositional phrase as postmodifier. There were eight versions of each of the beginnings half of which containing an animate head followed by an inanimate postmodifier and the other half the other way round. Half of the heads and postmodifiers matched in number and the other half did not. Also half of NPs were concrete whereas the other half were less concrete. The experiment was conducted in the same way as the first experiment and similar scoring categories were also used.

The results for the second experiment showed that 60.4 per cent ($n = 1236$) of all responses were correct, 2.3 per cent ($n = 48$) had agreement errors, 30.4 per cent ($n = 623$) had inflectional errors, and 6.9 per cent ($n = 141$) of the responses were

miscellaneous. Plural postmodifiers resulted in more errors than singular ones but there were slightly more errors when the postmodifier was animate regardless of the match or mismatch between the head. Also the mismatch between the head and postmodifier showed again more errors in agreement. Moreover, uninflected verb forms were more common after inanimate postmodifier than an animate one. This was explained by the fact that many more lexical verbs take an animate rather than an inanimate subject and as it was harder to find a proper verb for an inanimate noun, the participants used the copula *be* more often which then creates more opportunities for agreement errors. Also, when the postmodifier was concrete rather than abstract, more errors appeared but this was mostly the case with plural postmodifiers. Moreover, there were more repetition errors when there was a mismatch between the head and postmodifier and also when the head was plural.

As structural relations did not seem to be relevant to the agreement process, the third experiment focused on plurality and sentence position. The 64 participants were again undergraduates from the Michigan State University and they were given extra credits for taking part in the experiment. None of them had been in the previous experiments. The materials from the second experiment were used but all the prepositions were replaced with a relative pronoun *that*. This resulted in situations where there were two subjects in NP, the head and the subject in the relative clause, and there were also two verbs, the first one that agreed to the subject in the postmodifier and the second that agreed to the head, as in *The politician that the flag adorned was pleased with his reception* (taken from the experiment data). Again, half of the sentences had a mismatch between the head and postmodifier and the other half matched. Also the experiment procedures repeated the previous experiments. The scoring was mainly the same except that now the participants had to produce at least one complete clause to get their sentence scored correctly.

The results of the third experiment showed for the head nouns that 25.2 per cent ($n = 516$) of all responses were correct, 1.8 per cent ($n = 37$) had agreement

errors, 57.0 per cent (n = 1168) had uninflected verb forms, and 16.0 per cent (n = 327) of the responses were miscellaneous. Agreement errors occurred more likely when the head noun was animate and/or plural than being inanimate or singular, and also when there was a mismatch between the head and the noun in the postmodifier. The most agreement errors occurred when all these factors were present, i.e. when the head was animate, plural and mismatched with the postmodifier. All in all, plural heads brought in more errors than singular ones. Animate nouns also resulted in more errors with the inflection of the verbs, and especially plural nouns in the postmodifier induced more errors. Mismatch between the head and the postmodifier yielded also more reproduction errors than a matching situation. Moreover, there were more errors when the head was concrete than abstract but the differences were not significant.

The responses were scored also for the verbs in the postmodifying sentences and they resulted in somewhat different percentages. 59.9 per cent (n = 1227) of all responses in the postmodifiers were correct, 5.5 per cent (n = 113) had agreement errors, 11.6 per cent (n = 238) had uninflected verbs, and 23.0 per cent (n = 470) were miscellaneous. The results were generally similar to the results of the second experiment. There were more number errors when the head and the postmodifier mismatched and, again, plural nouns resulted in more errors than singular ones. Moreover, inanimate nouns yielded significantly more errors than animate ones. Also, the inanimacy of the noun in the postmodifier produced more errors to the main verb when there was a mismatch in animacy between the head and the noun in the postmodifier.

To sum up the results, the researchers concluded that, firstly, when there was a mismatch in number between the head and the postmodifier, errors occurred. The first experiment brought out more errors when the head was singular and the postmodifier plural than the other way. Moreover, the length of the postmodifier had no significant effects on the occurrence of errors. Secondly, as the second experiment showed, the animacy of the postmodifier did not have an effect on the occurrence of errors but, as the third experiment revealed, when

animacy of NP created problems in comprehending the meaning of the subject, errors occurred frequently. Thus, semantic features of the subject do not affect the agreement process unless they make understanding of NP more difficult. Thirdly, if plurality and position of NP would alone determine agreement, the results from the second and third experiment should have been similar; verbs should have agreed mostly with the head and occasionally with the noun in the postmodifier. However, the results were quite the opposite. All in all, the plurality of the postmodifier caused significantly more agreement errors than a singular postmodifier and this was shown in all three experiments. Consequently, this would suggest that the syntactic features have a bigger role in the agreement process and also the complexity of NP complicates performance.

Accuracy in oral and written production was studied by Ruin (1996). The participants were 72 first-year English students in the Uppsala University with a minimum grade of 3 on a 5-point grade scale. The participants were divided into groups of 20 to 25 students and they had a 90 minute grammar course altogether 17 times during nine weeks period. As material they had a school grammar and an exercise booklet of translation sentences. The class time was used to explain points of grammar and to discuss translation alternatives and answer the participants' questions. The purpose of the course was to improve the students' accuracy orally and in writing and their written performance was tested in a final exam. The data consisted of three different written pre-tests and post-tests: fill-in and sentence translation, translation from Swedish to English, and a composition. She calculated the error ratios by computing errors per words.

Ruin found out that the participants had the biggest difficulties in the number of the subject, i.e. sentences with complex NPs and relative clauses had more agreement errors than sentences with simpler NPs as a subject, or when no relative clauses were included. She added that if there was a contrast between the verb form expressing the singular/plural and the third person/other

person, an error was more likely to occur than if it was a clear case. Also, when the singularity or plurality of the subject was uncertain or ambiguous, errors were very common. For example, the 3rd person singular *-s* was mastered better when the subject was a simple noun or a pronoun and was followed immediately by the verb. Moreover, few agreement errors were made when the copula *be* was present. According to Ruin, this could be explained partly by the fact that the participants try to avoid difficult structures where they are more likely to make errors and thus they produce structurally simpler language. The participants also had the fewest errors in a composition which might be due to the same reason.

Ruin concluded that there were three causes to subject-verb agreement errors. Firstly, the participants either forgot to add the 3rd person singular *-s* to the verb or they added it to the verb after a plural subject. Secondly, the participants sometimes misinterpreted the number of the subject with countable and uncountable nouns and this was especially the case when countable and uncountable nouns differed from their L1. Thirdly, the participants had difficulties in identifying the number of the subject when there was a bigger distance between the subject head and the verb. According to Ruin, errors with complex NPs as a subject were not caused by transfer from L1 but were simply due to the lack of ability to cope with complex linguistic structures. The results could also depend on the participants' level of language knowledge at a certain point of their development.

The reasons behind the omission of verbal inflection in L2 acquisition were studied by Ionin and Wexler (2002). As hypotheses they suggested that, firstly, L2 learners would produce non-finite forms in the place of finite forms. Secondly, they would have no problems with abstract categories and so there would be little or no incorrect finiteness inflection. Thirdly, they also expected to find more *be* forms than forms with inflectional affixes.

The participants were 20 Russian children acquiring English as L2. The participants were aged between 3 and 13 years. Ten of the participants had lived in the USA for less than a year before participating the study, six had lived in the USA longer than one year but less than three years, and four of the participants had been born to Russian-speaking families in the USA or Canada and had been exposed to English three years or less before the experimentation. Of the 16 participants not born in the USA, only 3 had had exposure to English prior to their arrival to the USA. At the time of the study, all participants could speak and understand English. 15 of the participants were attending school and only one of them received special help with English.

English speech samples were collected from all of the participants in individual conversations by the researchers in which the participants were asked to talk about their friends or schoolwork, or to describe pictures in storybooks. The conversations lasted approximately 30 to 60 minutes and they were recorded for later transcription. A second sample was collected two to five months later from eight participants. The use of verbal inflection of four types of morphemes (the 3rd person *-s*, the past tense *-ed*, the *be* auxiliary and the *be* copula) was studied to see how often the participants failed to use the inflection correctly in obligatory occasions. They also examined the number of tense/agreement errors in the data which included the use of a *be* form for inappropriate person, number of tense and *-s* used with other tenses than the 3rd person singular.

The results showed that the omission of inflection was high with all types of morphemes but omission of morpheme inflection was higher for inflectional affixes than for *be* forms as they had hypothesised. The past tense *-ed* was not used incorrectly in a nonpast context but an irregular form was inflected incorrectly with *-ed* morpheme a few times. There were also a few cases where the 3rd person singular and *be* forms were inflected incorrectly. All in all, there were very few tense/agreement errors in the data. The participants also used *be* forms much more frequently than inflectional affixes. The low use of affixial inflection might be due to the fact that the affixial status of *-s* and *-ed* makes

them more difficult to acquire than the inflection of *be* forms. They also discovered that the participants overgeneralized the *be* forms as they used *be* forms with uninflected thematic verbs in utterances where a progressive VP would have been needed, such as in *the lion is go down*. This phenomenon could be found in over half of the transcriptions. However, when *be* was used, in most of the cases it was not inflected but was followed by an inflected main verb.

Although the participants produced high rates of uninflected verbs, they almost never used agreement morphemes for inappropriate tense, person or number. They were also more proficient in the use of *be* forms than in the use of main verb agreement but this did not seem to be caused by problems in affixation. According to the researchers, this might be due to the different ways on how these two types of verbs function in sentences. As auxiliaries and copula raise to tense in English, main verbs can be located further away from NP which might not always lead L2 learners to correct morphological agreement. Thus morphological inflection with *-s* and *-ed* might be more difficult to acquire than *be* forms that always raise to tense. Moreover, child L2 learners need time to acquire language-specific morphological rules and it is easier for them to learn that *be* forms need to be inflected always whereas the unraised lexical verbs have their own rules of inflection and might even consider *-s* and *-ed* inflection optional before internalising the rule. The participants may have even used *be* forms as a substitute for morphological inflection or to mark where inflection is supposed to be which would explain the overgeneralization of the *be* forms.

The variability to the overuse of 3rd person singular -s by Japanese learners of English and the possible reasons behind this variation was investigated by Shibuya and Wakabayashi (2008). They sought to find out whether the way in which the number feature of NP is marked influences the sensitivity to subject-verb agreement.

The participants were twenty Japanese undergraduate learners of English at the University of Chuo and they had all started studying English around the age of

12. Their English skills were measured with the Oxford placement test prior to the research for the study. They were assessed to be on the intermediate level. Nine native speakers of English also participated the study. A self-paced word-by-word reading task was used where the reading time of each word was recorded. The reading time of grammatical sentences was compared to the reading time of ungrammatical sentences to see whether the time spent for the ungrammatical sentences was longer which would indicate the participant to be sensitive to the ungrammaticality of the sentence. The experiment was done individually in front of a computer and the session took 25 to 35 minutes. After reading the sentence a yes/no comprehension question about the sentence was presented to the participants to make sure they had understood the sentence. Then a yes/no grammaticality judgement question appeared on the screen. Response times to all these were recorded. The experiment material consisted of five types of sentences: 2nd person, Proper noun and Proper noun, Simple plural subject, *These* + quantifier, and 3rd person singular. Altogether 40 sentences were present, half of them being grammatical and the other half ungrammatical.

Three hypotheses were presented. Firstly, Japanese learners of English would be insensitive to the omission of 3rd person singular *-s*. Secondly, the learners would be insensitive to the overuse of 3rd person singular *-s* if it violates the person agreement. The third hypothesis stated that the learners would be more sensitive to the overuse of the 3rd person singular *-s* when two proper nouns are connected with *and* or when *These* + quantifier functions as NP but not when a simple plural subject is as NP.

The results showed that the Japanese learners of English were actually more accurate with judging the grammaticality of the sentences than the native speakers were. Also accuracy with 2nd person as NP was higher than accuracy with 3rd person singular as NP in both groups. However, 3rd person singular as NP caused more difficulty for the Japanese learners than any other types of NPs. Moreover, reading times for ungrammatical sentences showed clear

delays for the native speakers with all types of NPs which suggests that they were sensitive to the ungrammaticality of the sentences. The Japanese learners showed longer reading times for ungrammatical sentences where NP was either 2nd person, two proper nouns, or *These* + quantifier. When NP was a simple plural subject, no significant differences in the reading times of grammatical or ungrammatical sentences appeared. But when 3rd person singular functioned as NP the reading times for grammatical sentences were actually longer than for ungrammatical ones.

In sum, the Japanese learners were sensitive to person disagreement as there were significant differences in the reading times of grammatical and ungrammatical sentences and they also noticed the overuse of 3rd person singular when reading the ungrammatical sentences. However, the learners were not sensitive to the omission of 3rd person singular *-s* in ungrammatical sentences with 3rd person singular as NP which supports the first hypothesis presented above. Bound morphemes caused insensitivity to the overuse of 3rd person singular with a plural NP but with free morphemes or a conjunction (*and*) the learners could judge the grammatically correctly. Native speakers turned out to be sensitive to both overuse and omission of 3rd person singular *-s*. Thus the results might be more due to the fact that the learners are not sensitive to the number feature of NPs than being aware of the features of the verb.

The results of the previous studies were interesting as they showed what types of problems L2 learners of English usually have with subject-verb agreement. However, the previous studies regarded the phenomenon and its features as such and did not take into account the proficiency level of the learners. Some of the studies had native speakers as a control group whose results were contrasted with those of the participants but the L2 learners were still studied as one group and they were not differentiated according to their skills in TL. This was also demanded by Ellis and Barkhuizen (2005: 79) as they suggested that learners should be studied at the same stage of development (see section 2.2.2).

LL can, however, vary greatly depending on the level of the learners but there can also be individual variance. Therefore the present study will take into account the L2 learners' proficiency in English as it examines learner productions that have been divided into different levels according to the learners' language proficiency. The present study will also take into account the overuse of the 3rd person singular -s whose absence was also criticised earlier (see section 2.2.2). The present study and its methodology are presented in the next section.

4 THE PRESENT STUDY

From now on the focus will turn to the present study by outlining design of the research. Firstly, the chapter presents the motivation for the present study and states the research questions. Secondly it will introduce the data and its collection in more detail and finally also summarize the theoretical framework used for analysis.

4.1 Aims of the study

Subject-verb agreement has not been studied in Finland lately even though it is one of the basic features of English grammar. The previous studies on the topic have not taken into account the proficiency scales of the CEFR, as it is a fairly new publication, and have mostly been conducted elsewhere with different L1s than Finnish; therefore the present study will shed light on the phenomenon from a slightly new perspective. As the present study is also a part of the CEFLING project, its purpose is to analyse development on a particular area of language proficiency as compared to the proficiency scales. The other subprojects under the project have studied various other linguistic features (see the project's web pages) and the present study presents new information based on the data collected for the project. Therefore the research questions to this study are:

- 1) What is the relationship between learners' use of subject-verb agreement of finite verbs in the present tense and their proficiency level assessed by the CEFR scales?
 - a) Does the frequency of use vary from one proficiency level to another?
 - b) How accurately can they use the feature?

2) What is the relationship between learners' use of subject-verb agreement of finite verbs in the present tense and the complexity of the noun phrase?

- a) Does the type of the noun phrase have an influence on the accurate use of the feature?
- b) Are there differences from one proficiency level to another in using different nouns?

Due to the problems and limitations of the previous studies in describing the usage of learner language (see sections 2.2 and 2.3 for details) the present study aims at taking into consideration both the learners' proficiency level and also the complexity of their language use. This time the focus on complexity is with subjects. The study also wants to find out how well the learners master subject-verb agreement of finite verbs in the present tense in their written productions and whether there are differences in the mastery between different proficiency levels and how the feature develops through these levels. The present tense of the finite verb, i.e. the main verbs and the copula, was chosen as it shows the agreement more clearly than other tenses. Also modal auxiliaries were excluded as they do not show the number of the subject. All the aspects of the CAF triad presented above (see section 2.3.1) are taken into consideration and compared between the different stages. In addition, the study aims at finding out whether there is a relationship between the complexity of NP and the correct use of verb-subject agreement on different levels of language development. Since learner language is supposed to get better with higher proficiency, it could be hypothesised that learners on higher levels master the particular feature better than learners on lower levels and their NPs should also get more complex, and there should also be differences in the results. Therefore the accuracy of different types of NPs is compared to see whether the accuracy also follows development.

In order to find answers to the research questions above, written data produced by seventh and ninth graders are analysed in terms of fluency, accuracy and complexity. The next section will explain the data collection as well as introduce the participants of the present study.

4.2 Data collection

The data for the present study consists of written tasks from Finnish secondary school (grades seven and nine) pupils that have been collected for the Cefling project. The participants were aged from 13 to 16 and had started their English studies in third grade. The data for the Cefling project consists of a total of 527 texts that were gathered during regular English lessons. There were five different written tasks: a message to a friend, a message to a teacher, an e-mail to a net store, giving an opinion, and telling about the most exciting or frightening day of their lives. The data has then been evaluated by four different evaluators making use of the descriptive levels in CEFR. In order to place a production on one of the proficiency levels, three of the four evaluators had to agree on the level. The data was also assessed to range from level A1 to level B2 as no C-scaled productions were found. The present study does not look into possible development from grade seven to nine as the grades were not taken into consideration but concentrates on comparing differences between different levels of proficiency.

As all sentences include subject-verb agreement, the number of objects would have been too enormous for the present study had all the tasks been included. Therefore the present study consists of 141 texts produced for the Task 4 and the distribution of the texts on different levels is shown in Table 1.

Table 1. The number of texts on different levels

Level	Texts
A1	45
A2	60
B1	34
B2	2

45 were assessed to be level A1 texts, 60 level A2 texts, 34 level B1 texts, and only 2 reached level B2. The learners were asked to give an opinion on one of the topics: 'Boys and girls should go to different classes at school' or 'No mobile phones at school!', and also to give reasons for their arguments. The assignment description is in Finnish below.

<p>T4 Mieliopide</p> <p>Valitse aiheista yksi ja kirjoita, mitä ajattelet aiheesta. Perustele mielipiteesi.</p> <ol style="list-style-type: none"> 1. Boys and girls should go to different classes at school 2. No mobile phones at school!

Task 4 was chosen not only because it included most present tense forms but also because it controlled learner production the least whereas the other tasks had more precise instructions on what the text should include.

4.3 Analysis of the data

In analysing the data the Cefling project has been using a DEMfad model (The DEMfad Model, Franceschina et al. 2006), which combines different dimensions of the data and is therefore used to track language development comparable across the levels, domains and languages. The model (Figure 3) is based on

three different concepts in language development, namely fluency, accuracy and complexity (see section 2.3 for details).

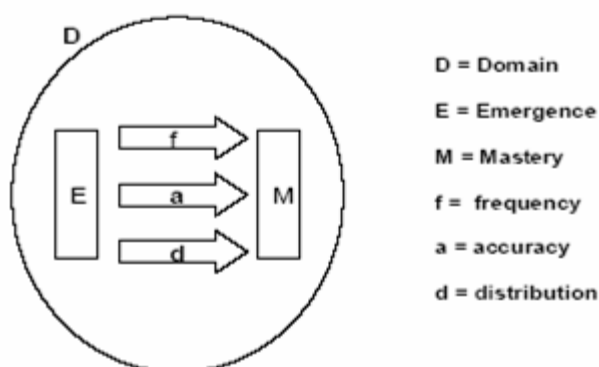


Figure 3. The DEMfad model

The domains (D) are the areas of developing language skills, be it, for example, a certain construction, a set of vocabulary or a certain linguistic device. The emergence (E) refers to the first occurrence of a certain domain, for example, the use of the 3rd person singular *-s*, and it is expected to emerge at a certain stage in language development. Mastery (M) is defined as approximately target-like use of the domain which has been set on average at a tentative level of 80% correctness (see Martin, Mustonen, Reiman and Seilonen 2010). There are, however, many stages between emergence and mastery of a linguistic feature, and the development can be multifaceted. Frequency (f) is related to the concept of fluency, productivity being one of its measures, and it means how often the observed feature occurs. It is calculated per 1000 words of the learners' production. Accuracy (a) indicates how often a linguistic feature is supplied correctly and it is always compared to TL, whereas fluency can be defined without referring to it. The term distribution (d) was chosen to cover several features of learner language development some of which concerned complexity and some concerned variability. Frequency and accuracy can be analysed by the means of a quantitative analysis but distribution needs a more qualitative approach to data analysis.

In the present study the model is defined as follows. The domain of the study is subject-verb agreement of finite verbs in the present tense in written productions, and mastery of the feature follows the same 80 % limit used in the DEMfad model. Frequency, on its part, means the extent of use of subject-verb agreement per 1000 words, and it is calculated by looking at the agreement as a phenomenon. All cases of the phenomenon at a certain descriptive scale are counted and the figures are calculated per 1000 words. Accuracy stands for the percentage of how often subject-verb agreement in each case is produced correctly. When using a language, obligatory occasions are created when, for example, starting a sentence with a third person subject, an obligatory context is created for the 3rd person *-s*. Therefore the analysis of accuracy bases on obligatory contexts: all the sentences are supposed to include the phenomenon at issue, and as in English the subject defines the number of the verb, all cases are examined and compared to TL. The percentage of correct use is then calculated and a comparison between the different levels is made. Complexity is measured by looking at NPs: the accuracy in subject-verb agreement of different types of NPs is compared to see if some NPs induce more errors than others. The present study does not commit to emergency as such as the phenomenon at issue is supposed to be found in every clause and is one of the bare bones of English sentence structures.

Having introduced the data and the methods of the present study, next chapter will present the results of the study. After that the focus will move on to discussing the results in more detail.

5 FINDINGS

This chapter reports the findings of the present study. There were some cases that were left out from the analysis as they lacked an important grammatical feature. There was one case on level A1 that lacked a subject and three cases that lacked a verb. On level A2 three cases lacked a subject and another three cases lacked a verb. Moreover, there was one case on level A1 where the subject was written in Finnish and another case on level A2 where the verb was written in Finnish. These above cases were not analysed as it could not be interpreted what forms the learners had tried to produce. Section 4.1 presents the figures on frequency on the use of subject-verb agreement. Section 4.2, for its part, shows the figures on accuracy. The results on distribution are presented in section 4.3 by looking at different NPs the learners have used and comparing their accuracy rates on using these different NPs from one level to another.

5.1 Frequency of subject-verb agreement in the data

The first research question was about the relationship between the learners' use of subject-verb agreement and their proficiency levels and the question was divided into two sub questions. In order to answer the first part of the research question about the frequency of use of subject-verb agreement and the possible differences from one level to another, a comparison between the levels was made to find out how often learners on each level used the feature in their texts. The levels clearly had a different number of texts as 60 of them were assessed to reach level A2 and only 2 of them level B2. The texts themselves showed also differences in length and also in the use of the subject-verb agreement from one level to another. The comparison for use of subject-verb agreement and also the frequency can be found in Table 2.

Table 2. Frequency counts on subject-verb agreement on different levels

Level	Texts	Wordcount	N	Frequency
A1	45	1600	200	125
A2	60	4177	418	100
B1	34	3693	318	86
B2	2	307	23	75

The total word count on level A1 was altogether 1600 words. On level A2 it was 4177 words, on B1 3693, and on B2 307 as there were only two texts on this level. There were 200 cases of subject-verb agreement on level A1, and this resulted in a frequency of 125 words per 1000 words on the use of the phenomenon. Level A2 had 418 cases which resulted in 100 words frequency per 1000 words. 318 cases were found on level B1 resulting in 86 words per 1000 words frequency whereas on level B2 the frequency was 75 words per 1000 words with 23 cases. Subject-verb agreement was used on the average most frequently on level A1 and least frequently on level B2 so the learners on the lowest level were the most eager to use the feature in their texts.

The texts varied considerably in their lengths at different levels ranging from six to 204 words per texts. The average number of words per text is shown in Table 3.

Table 3. The average number of words per production

Scale	Average number of words per text
A1	35.6
A2	69.6
B1	108.6
B2	153.5

On level A1 the mean value was 35.6 words per production. On level A2 it was 69.6, on B1 108.6, and on B2 the mean value was 153.5 words per production.

The average number of words per production rose quite steadily from level to level. The second part of the question will be answered in the next section.

5.2 Accuracy in the use of subject-verb agreement in the data

The second part of the first research question sought to find out how accurately the learners were able to use subject-verb agreement in their texts. Hence accuracy was measured by comparing the number of correctly used cases of subject-verb agreement to the number of the total amount of cases per different proficiency levels. Table 4 presents the numbers for accuracy on different proficiency levels.

Table 4. Accuracy counts on different levels

Levels	N	N correct use	N incorrect use	Accuracy (%)
A1	200	169	31	84.5
A2	418	374	44	89.5
B1	318	304	14	95.6
B2	23	23	0	100.0

Out of 200 cases of subject-verb agreement on level A1 169 were produced correctly which resulted in 84.5 per cent accuracy. On level A2 the feature was produced correctly in 374 cases out of the total number of 418 cases. This resulted in 89.5 per cent accuracy. On level B1 accuracy was 95.6 per cent as the number of correct use was 304 out of 318 cases. All 23 cases were produced correctly on level B2 the accuracy being 100 per cent. All levels showed fairly high numbers of accuracy, the lowest being 84.5 per cent and the highest 100 per cent.

The levels were also compared according to the average number of cases per texts so it could be seen if there were differences in the use of the feature between the levels. This comparison is shown in Table 5.

Table 5. The number of cases per text on different levels

Scales	Number of cases per texts
A1	4.4
A2	7.0
B1	9.4
B2	11.5

On average, the number of cases per text also rose from level to level: A1 had 4.4 cases, A2 7.0 cases, B1 9.4 cases, and B2 11.5 cases per text. Even though there were most cases on level A2 and least cases on level B2 (see Table 4), the learners on level B2 had used the feature more frequently than the learners on any other level.

5.3 Distribution of different types of noun phrases in the data

The second research question compared the learners' use of subject-verb agreement to the complexity of NP and was divided into two subquestions which will be answered below. NPs were defined according to their functions. The division of NPs into subgroups with a few examples is shown below according to what types of NPs were found in the data. The division was made according to the division in Downing and Locke (2006, modules 45-50).

1. Personal pronouns (*I, you, he*)
2. Indefinite pronouns (*somebody, anyone, nothing*)
3. Demonstrative pronouns (*that, this, those*)
4. Possessive pronouns (*my, your, our*)
5. Quantifiers (*some, every, any*)
6. Classifying genitives (*John's car, dog's toy*)
7. Unstressed there (*there are many boys*)
8. Article + head (*the house, a cat*)
9. Adjective phrase + head (*great day*)
10. Head + prepositional phrase
11. Head + adjective phrase
12. Dummy it (*it rains*)
13. Relative clause (*what, who, which*)
14. Whole clause
15. Other nouns (*mobile phone*)

Of NPs classified above, the subgroups 3, 4, 5, 6, 8, and 13 belong to determining elements. Determining elements are such that particularise the noun from other surrounding elements. The use of determiners tells how the noun that it refers to can be identified. For example, when a definite article is used, the noun can be identified either in the text, in the situation or in common knowledge of the world, such as *the sun*. Groups 9, 10, 11, 14, and 15 are modifying elements that divide further into premodifiers and postmodifiers. Groups 9, 14 and 15 are premodifiers, and groups 10 and 11 belong to postmodifiers. Modifiers describe the characteristics or qualities of the head noun which can be either permanent or non-permanent. Premodifiers usually express the first one with inherent, permanent or characteristic features of the head noun and are thus placed before the head to give the status of the features, such as *my car*. Postmodifiers express the latter one with changing features of the noun.

The first part of the second research question aimed at finding out whether the different types of NPs would have an influence on the accurate use of subject-verb agreement. The purpose of the second part of the question was to compare the different proficiency levels to find out whether they varied in their accuracy with using the feature. Both questions will be answered simultaneously. The

total accuracy on level A1 was 84.5 per cent but it divided differently among different subjects, and some subjects were more popular than others. Table 6 presents how the use of different nouns divided between NPs and how the accuracy differed between the subjects on this level.

Table 6. Accuracy on different types of nouns on level A1

Type of NPs	N	N correct use of NPs	N incorrect use of NPs	Accuracy (%)
Personal pronouns	109	106	3	97.2
Indefinite pronouns	17	6	11	35.3
Demonstrative pronouns	5	5	0	100.0
Possessive pronouns	8	6	2	75.0
Quantifiers	5	3	2	60.0
Unstressed <i>there</i>	1	1	0	100.0
Article + head	41	33	8	80.1
Adjective phrase + head	1	1	0	100.0
Head + prepositional phrase	2	0	2	0.0
Dummy <i>it</i>	8	8	0	100.0
Other noun	14	10	4	71.4

When the subject was a personal pronoun, accuracy reached a percentage of 97.2 with 109 cases but with 17 cases of indefinite pronouns it was only 35.3. With demonstrative pronouns accuracy was 100 per cent with 5 cases as well as with an adjective phrase and a head, dummy *it* and with the unstressed *there*. An adjective phrase was used once, dummy *it* 8 times and the unstressed *there* one time. With eight cases of possessive pronouns accuracy reached 75.0 per cent, and quantifiers resulted in 60.0 per cent accuracy with 5 cases. The article system and a head reached a result of 80.1 per cent accuracy, and if the head was accompanied with another noun, the percentage for accuracy was 71.4. The article system was used most commonly with 41 cases, and another noun preceded the head 14 times. The most common other noun was *mobile* as the topic was about using mobile phones at school (see section 4.2). A head and a prepositional phrase as a postmodifier resulted in zero per cent accuracy. As

the level of mastery in the DEMfad model was 80 per cent, it could be said that personal pronouns, demonstrative pronouns, unstressed *there*, article and the head, adjective phrase and the head, and dummy *it* were mastered by the participants on this scale. Due to the fairly small amount of cases it cannot be generalized whether the more complex NPs would cause more problems in producing subject-verb agreement correctly. However, according to the findings the most frequently used NPs, i.e. personal pronouns and article and the head, were mastered whereas with the less frequent ones mastery was more or less variable.

Level A2 had a total accuracy of 89.5 per cent that also varied between different subjects. The results can be found in Table 7.

Table 7. Accuracy on different nouns on level A2

Types of nouns	N	N correct use of NPs	N incorrect use of NPs	accuracy (%)
Personal pronouns	208	197	11	94.7
Indefinite pronouns	28	18	10	64.3
Demonstrative pronouns	15	14	1	93.3
Possessive pronouns	6	4	2	66.7
Quantifier	1	1	0	100.0
Classifying genitive	2	1	1	50.0
Unstressed there	15	12	3	80.0
Article + head	94	81	13	86.2
Adjective phrase + head	3	2	1	66.7
Dummy <i>it</i>	36	36	0	100.0
Relative pronouns	8	6	2	75.0
Whole clause	1	1	0	100.0
Other noun	16	13	3	81.3

A quantifier as a determiner; the dummy *it* and a whole clause as a postmodifier all produced a 100 per cent accuracy. A quantifier was used once as well as the whole clause whereas the dummy *it* reached 36 cases. Personal pronouns reached 94.7 per cent with 208 cases and demonstrative pronouns 93.3 per cent with 15 cases. Indefinite pronouns resulted in a slightly lower

number of 64.3 per cent with 28 cases as also did the 6 cases of possessive pronouns with 66.7 per cent whereas relative pronouns scored higher with 75.0 per cent which were used eight times. The article system with the head was used altogether 94 times with 86.2 per cent accuracy while another noun with the head reached fairly similar numbers of 81.3 per cent but was used only 16 times. As a determiner a classifying genitive reached 50 per cent accuracy, and an adjective phrase as a premodifier scored 66.7 per cent. The first was used twice and the latter three times. The unstressed *there* with 15 cases resulted in 80 per cent accuracy. On this level personal pronouns, demonstrative pronouns, quantifiers, unstressed *there*, article and the head, dummy *it*, a whole clause as a postdeterminer, and other noun were mastered according to the DEMfad model. On this level all but one more frequently produced NPs reached the level of mastery with a correctly produced subject-verb agreement. The indefinite pronouns turned out to be challenging as well as on level A1. The dummy *it* was produced fairly frequently and reached a 100 per cent level of mastery.

The total accuracy for level B1 was 95.6 per cent. The results are presented in Table 8.

Table 8. Accuracy on different nouns on level B1

Type of noun	N	N correct use of NPs	N incorrect use of NPs	Accuracy (%)
Personal pronouns	138	137	1	99.3
Indefinite pronouns	20	18	2	90.0
Demonstrative pronouns	17	17	0	100.0
Possessive pronouns	2	2	0	100.0
Quantifier	11	11	0	100.0
Classifying genitive	1	1	0	100.0
Unstressed there	15	10	5	66.7
Article + head	61	56	5	91.8
Adjective phrase + head	4	4	0	100.0
Head+ prepositional phrase	2	2	0	100.0
Dummy it	42	42	0	100.0
Relative pronouns	3	3	0	100.0
Whole clause	2	2	0	100.0
Other noun	10	10	0	100.0

Of the 14 different types of nouns, 10 reached a full 100 per cent accuracy. Of these nouns demonstrative pronouns were used 17 times, quantifiers 11 times and another noun 10 times. The rest of the different types received smaller numbers. An adjective phrase as a premodifier was used four times, and a relative pronoun three times. A possessive pronoun, a prepositional phrase as a postmodifier, and a whole clause as a premodifier were all used twice. A classifying genitive had only one case. Personal pronouns with 138 cases did almost as well with 99.3 per cent accuracy whereas the unstressed *there* with 15 cases was left behind with 66.7 per cent accuracy. Indefinite pronouns resulted in 90.0 per cent accuracy and were used 20 times. The article system with the head scored 91.8 per cent with 61 cases. All the other types of nouns except the unstressed *there* were mastered on this scale. The learners on this level are starting to reach a complete mastery of subject-verb agreement.

Level B2 reached a total 100 per cent accuracy. The numbers are shown in Table 9.

Table 9. Accuracy on different nouns on level B2

Type of noun	N	N correct use of NPs	N incorrect use of NPs	Accuracy (%)
Personal pronoun	7	7	0	100
Demonstrative pronoun	2	2	0	100
Possessive pronoun	1	1	0	100
Unstressed there	3	3	0	100
Article + head	7	7	0	100
Dummy it	2	2	0	100
Relative pronoun	1	1	0	100
Other noun	1	1	0	100

It had 23 cases of the subject-verb agreement which divided between eight different types of nouns: personal pronouns (7 cases), demonstrative pronouns (2 cases), possessive pronouns (1 case), unstressed *there* (3 cases), article system and the head (7 cases), dummy *it* (2 cases), relative pronouns (1 case), and other noun plus the head (1 case). All the nouns used in the productions were mastered according to the 80 per cent criteria of mastery in the DEMfad model but the relatively small number of finite verbs in the present tense might be due to the learners' language proficiency which allows them to use even more complex language.

Accuracy varied depending on whether the subject was accompanied by a determiner or a modifier. The comparison is presented in Table 10.

Table 10. Accuracy counts on determiners and modifiers.

	A1		A2		B1		B2	
	N	%	N	%	N	%	N	%
Determiner	60	78.3	126	84.9	95	94.7	11	100.0
Modifier	17	35.3	20	80.0	18	100.0	1	100.0
Premodifier	15	73.3	20	80.0	2	100.0	1	100.0
Postmodifier	2	0.0	-	0.0	16	100.0	-	0.0

On level A1 a determiner was used 60 times and a modifier 17 times. The accuracy was 78.3 per cent when the subject was preceded by a determiner but 35.3 per cent when there was a modifier with a noun. A postmodifier resulted in zero per cent accuracy with two cases, whereas a premodifier showed 73.3 per cent accuracy with 15 cases. Level A2 revealed 84.9 per cent accuracy when a determiner was used, and with a modifier the percentage was 80.0 with all being premodifiers. A determiner was used 126 times and a modifier 20 times. On level B1 a determiner was used 95 times and the accuracy was 94.7 per cent and when either a pre- or a postmodifier was used, the accuracy was 100 per cent with 18 cases. Level B2 reached a total of 100 per cent accuracy with both a determiner and a modifier. A determiner was used 11 times and a premodifier once. Altogether, a determiner with NP was mastered more often than NP accompanied by a modifier.

The accuracy varied slightly also depending on the singularity or plurality of the noun. This is shown in Table 11.

Table 11. The number of cases and accuracy on singular and plural nouns

	A1		A2		B1		B2	
	N	%	N	%	N	%	N	%
Singular	149	82.6	278	91.0	218	98.2	15	100.0
Plural	51	84.3	139	85.6	100	91.0	8	100.0

A singular subject was used 149 times and resulted in 82.6 per cent accuracy on level A1 whereas with a plural form with 51 cases, the accuracy was 84.3 per cent. Level A2 showed accuracy figures of 91.0 per cent and 85.6 per cent for singular and plural subjects, respectively. A singular noun was used 278 times and a plural noun 139 times. On level B1 the accuracy for singular subjects was 98.2 per cent with 218 cases and for plural subjects with 100 cases it was 91.0 per cent. As the accuracy figure for level B2 was 100 per cent, there were no differences between singular and plural forms – a singular noun was used 15 times and a plural noun eight times. Singular nouns were used more frequently than plural nouns and they also reached a slightly higher level of mastery. All in all, there were no significant differences in the use of singular or plural nouns in terms of subject-verb agreement.

The present study found quite interesting results on subject-verb agreement across the levels determined by the CEFR. All the findings are discussed in more details in the following chapter.

6 DISCUSSION

As the study included only finite verbs in the present tense, quite large parts of the productions had to be left out from the analysis as other verb forms than the finite verb in the present tense were used to convey messages. There were even some texts that were left out completely as they had zero cases with finite verbs in the present tense. On level A1, on the average, most sentences were written in the present tense with a finite verb, but on higher levels, for example, the past tense and auxiliaries were used in growing numbers which could not then be taken into consideration when analysing the data. In general the texts got more complex with higher levels leaving out more and more possible cases. This chapter summarizes and compares the results of the present study to these of the previous studies.

The first research question was about frequency and accuracy in the use of subject-verb agreement with finite verbs in the present tense and how this was related to the learners' proficiency levels as assessed by the CEFR scale. It was assumed that the learners' language skills would get better on higher levels and with this they would also become more fluent and accurate. However, surprisingly, frequency dropped while the learners' language skills got better. On level A1 the average number of words per text was the lowest whereas level B2 it was the highest. On level A1 the average number of words per text was 35.6 with a frequency of 125 words per 1000 words. On level B2 the corresponding figures were 153.5 and 75, respectively. The frequency numbers are notably lower on level B2 even though the average number of words per text is more than four times higher. This could be explained at least partly by the fact that the texts were more complex on level B2 with more use of, for example, auxiliaries, infinitive forms, and subject complements. Texts on level A1 had simpler sentences with fewer numbers of subordinate clauses or other complex sentence structures. Already on level A2 some improvement could be noticed in the use of more complex sentence structures, which could be seen in

the average number of words per text – the number almost doubled from 35.6 on level A1 to 69.6 on the level A2. Frequency dropped to 100 per 1000 words even though level A2 had more texts as well as cases than other levels. On level B1 the average number of words per production was already 108.6 with a frequency ratio of 86 per 1000 words. When jumping from level A2 to B1, the differences in sentence structures was even clearer than between levels A1 and A2. However, frequency did not drop in the same relation on level B2 as it did between levels A1 and A2.

Even though frequency got lower from level to level while the number of words per text rose, the number of cases per text got higher when moving to higher levels of proficiency. However, also accuracy rose from level to level reaching a total 100 per cent on level B2. The higher the learners' texts were scored on the levels, the more accurate they actually became. Level A2 had the highest number of texts (see Table 1) as well as cases but the average number of cases was 7.0 whereas on level B1, for example, there were almost half the number of texts compared to level A2 and still the average number of cases per text was higher, 9.4. Accuracy rose from level to level, as was predicted, but it actually rose quite steadily as the biggest difference was 6.1 units between levels A2 and B1 and the smallest 4.4 units between levels B1 and B2.

The second research question focused on the complexity of NPs in relation to the use of the subject-verb agreement. It sought to find out whether different types of nouns would result in differences in accuracy and also whether there would be differences between the proficiency levels. As could be seen from the tables above, the higher the proficiency level was, the better accuracy also got. Surely there were some deviations from the ascending trajectory but all in all it could be said that the learners' language skills got better with higher ratings on the CEFR proficiency levels.

Personal pronouns are not very complex types of NPs but they brought some difficulties in all other levels except on level B2 where all cases were fully

mastered. The verb *be* did not turn out to be problematic even though it needs to be inflected separately for each person. There were some errors in the use of this verb with personal pronouns such as *they is* on level A2 or *I are* on level A1. These errors were not, however, very common whereas the verb *do* and regular verbs caused more problems to the learners. Errors such as *they doesn't meat*, *we needs* or *they all rings* occurred on levels A1, A2 and B1. Actually on one text on level A2 an *-s* was added to all verbs in seven out of ten cases and of these six of the subjects were plural, and of these four had *we* as a subject. This might be due to an overgeneralisation of the rule as might happen with the acquisition of some morphemes (Ellis and Barkhuizen 2005: 77). When learners acquire a particular morpheme, they start overgeneralizing this rule after first being able to use it correctly. In this case it might be possible that the learner has recently acquired the use of the 3rd person singular *-s* but is now overgeneralizing the rule to cover all verbs that follow a personal pronoun. There was, however, one case with the subject *we* that was produced correctly and it was when the subject was followed by the *be* verb, as in *we are*. The results were in line with Ruin's (1996) and Ionin and Wexler's (2002) studies as they also found more agreement errors when the copula was used after NP than when a regular verb was used. As learners overgeneralized the use of the copula in Ionin and Wexler's study, in the present study the learners overgeneralized the 3rd person *-s* with plural NPs. Accuracy figures did not rise from level to level as the lowest percentage was on level A2. On the other hand, level A2 had the most cases with personal pronouns and as there are more cases, there are also more opportunities for errors.

There were actually many texts where the subject-verb agreement was in some cases produced correctly with a certain type of a noun but in another case it was produced incorrectly in the same text. An example of this is from a level A2 text where the learner had used an indefinite pronoun incorrectly in two coordinated main clauses but in the very next clause the subject-verb agreement was produced correctly.

(1)...if someone call or write message secretly at class. Or if someone gets call or message.

In fact, the indefinite pronouns seemed to cause most of these types of problems in other levels as well. In another example from level A2 the participant had incorrectly used the subject-verb agreement with an indefinite pronoun when there was only the pronoun as a subject but in the previous sentence had managed to produce it right with a more complex subject including an indefinite pronoun plus a prepositional phrase.

(2)...if somebody in your family hurts himself or somebody attack you...

Different indefinite pronouns also misled the learners as with some pronouns they were able to produce the inflection correctly but with another type of a pronoun they produced it incorrectly.

(3)...everybody says diamonds are ... when someone send to me message...

The dummy *it* was mastered by 100 per cent on each level. It cannot, however, be concluded from the results whether the structure itself was easier to learn and therefore it would be easily mastered already at the low levels of language proficiency, or could this be an example of a memorized language chunk, as Ellis suggests (see section 2.3.1). Of all the different subjects produced in the texts, the dummy *it* always requires a singular form of the verb, and a chunk like *it is* is fairly easy to remember – it does not require too much from the working memory and is perhaps quite easily stored in the long-term memory as a chunk.

An interesting decline in the mastery of a certain type of a noun was with the unstressed *there*. Both levels A2 and B1 had 15 cases each but the accuracy was 80 per cent on the lower level when compared to only 66.7 per cent on the higher level. On level A2 the feature was mastered but on level B1, when the language proficiency was expected to be higher, the accuracy was quite far from the 80 per cent level of mastery. The unstressed *there* is of course more

difficult to learn as the entity it refers to comes after the verb and it also determines the number of the verb as in *There is somebody at the door* or in *There are three men at the door*. In the first case the entity *there* refers to a singular and thus the verb also needs to be singular whereas in the latter example the entity is plural and thus the verb also needs to be plural to agree the subject. The unstressed *there* cannot thus be learned as a chunk but learners actually need to acquire the rule to be able to produce it correctly. This does not, however, explain the decline of mastery on level B1.

Other noun followed by the head caused also more problems to some of the learners on levels A1 and A2 than the article plus the head. On level A1 the article with the head only just exceeded the level of mastery with 80.1 per cent but when there was another noun, the accuracy remained under the level of mastery with only 71.4 per cent. So in this case it could be said that when there was more than just an article and the head, it was more difficult for the learners to handle the inflection.

When NP was a more complex one, more errors occurred. Relative clauses as postmodifiers created problems on level A2 where mastery remained under 80 per cent. Level B1, for its part, showed a total mastery of the structure and on level A1 the learners had not produced any such structures. This was also the case with a head and a prepositional phrase as a postmodifier which could not be produced correctly at all on level A1 as accuracy was zero. On level B1 the structure was fully mastered with 100 per cent accuracy. The results are similar in this respect with Ruin's (1996) findings as she also discovered that the complexity of NP induced more agreement errors. It is, however, difficult to say whether these structures had just emerged in level A1 learners' language repertoire or if they have been introduced to the structures notably earlier as the productions were categorized only by the proficiency level and not by the grades the learners are on. Therefore it cannot be said if they were of the same age and were being taught the same things at the point of data gathering and would thus have similar prior knowledge for producing the structure, or if they

were of different age and grade and would thus have more knowledge and experience in their language use. A further examination should be made to compare learners of the same age to see how they differ from each other. This should then be done to learners of different age and grade and compare the results among and between the age groups.

Accuracy on determiners and modifiers improved along with language skills. The biggest differences were on level A1 where accuracy on using a determiner was 78.3 per cent but the accuracy on modifiers was only 35.3 per cent.

Determiners were over three times more frequently used than modifiers and they included simpler elements than the modifiers (see Table 9). Moreover, the postmodifiers on level A1 resulted in a zero per cent accuracy which could imply that the structure is more difficult and complex than a noun with a premodifier or a determiner. On levels A2 and B2 the structure was not used at all. On the other hand, on level B1 a postmodifier was eight times more frequently used than a premodifier but they were mastered equally well with 100 per cent accuracy on both. All in all, the use of a determiner was much more frequent on all the levels and also the accuracy was higher when a determiner was chosen.

What comes to the singularity and plurality of the nouns, the differences were not very significant on the different ability levels. On other levels, except on A1, the accuracy was higher with singular nouns than with plural nouns. Bock and Miller (1991) found similar results in all their three experiments; singular heads resulted in more correct responses than plural ones. However, the present study did not take a look at possible match or mismatch situations with the head and the postmodifier and, all in all, postmodifiers were very rarely used in the learners texts and would not thus have given reliable information on the occurrence of errors in such cases. Moreover, singular nouns were more frequently used than plural ones and thus perhaps also better acquired. On level A1 accuracy was higher when a plural noun was used but the singular noun was used almost three times more frequently with 51 and 149 cases,

respectively. However, the accuracy with both singular and plural nouns improved as the language skills of the learners got higher.

The findings of the present study showed some expected but also unexpected results. Moreover, as was discovered, different types of NPs also affected the correct use of subject-verb agreement. The present study was, however, rather a small scale study with quite a narrow data, so the results cannot therefore be generalized in a larger scale. The final chapter will discuss the present study all in all and assign the value judgement of the study. Also some suggestions for further study are also presented.

7 CONCLUSION

The aim of the present study was to examine how Finnish L2 learners of English succeed in using subject-verb agreement in their written productions and whether the results would show differences between different proficiency levels. There are various previous studies made on the topic too but they all have approached the feature only from a linguistic point of view ignoring the different language skills of the learners. Therefore the present study wanted also to take the aspect of language skills also into consideration. A further point of interest was also to look at different NPs to find out whether the complexity of the NPs would affect the correct use of subject-verb agreement. This final chapter will firstly discuss the main results of the present study and secondly the strengths and weaknesses of the study are discussed. Finally, some suggestions for further research on subject-verb agreement are also presented.

The present study revealed some fairly interesting results. Firstly, frequency on the use of subject-verb agreement dropped as the learners' skills got better because the learners started to use complex structures more as their skills got better. Secondly, accuracy rose along with the skills reaching 100 per cent by level B2. The overall accuracy was already fairly high on level A1 as it was 84.5 per cent. Hence the findings were in accordance with the hypothesis made earlier: the better and more accurate the learners' productions got, the higher they were ranked on the proficiency levels. Thirdly, more complex NPs caused more errors than simpler ones on the lower levels as on levels A1 and A2 the complex structures remained under the level of mastery. On level B1 the more complex structures were already mastered. Moreover, singular heads resulted in fewer errors than plural ones and they were also used more frequently. This might have been due to another finding which indicated that the copula *be* was mastered better than regular verbs as the regular verbs showed signs of overgeneralization; 3rd person singular *-s* was added also to verb forms that would have required a plural ending. This then caused incorrect agreement

with plural heads. Also, the dummy *it* was mastered 100 per cent correctly on all levels, which could indicate that it has been acquired as a chunk as it always takes the same form in the present tense (*it is*). In addition, the acquisition of subject-verb agreement seemed still to be in process as in some sentences the feature was produced correctly but in the very next sentence, or in the previous one, it was produced incorrectly.

The present study managed, firstly, to successfully combine examining learners' language use to their language proficiency by examining the fluency, accuracy and complexity in their productions and comparing the results between the different levels of proficiency. Subject-verb agreement has not been studied before with an emphasis on the learners' language skills and therefore the present study brought out new information into the field. Secondly the results also showed what types of NPs made it easier for the learners to manage subject-verb agreement and which NPs were more complicated for them to handle. The previous studies have examined the features of NPs from various points of views but they have not compared the results to the language skills of the learners. Moreover, the results showed how Finnish learners of English mastered the feature. The mastery of subject-verb agreement of Finnish L2 learners of English has not received much attention and so the results of the present study brought out updated information on the feature. Also the phenomenon has not been contrasted before to the proficiency level defined by the CEFR and therefore the results can be contrasted to other research under the Cefling project.

The sample was, however, relatively small and the results cannot thus be generalized in a larger scale. A further research with a larger pool of data would provide more comprehensible results on learners' language development. For example, among the texts there were only two that were rated to be on level B2 and therefore the results of this level are not completely reliable, or comparable to the other levels. The present study also examined the feature in only one type of task, and as was mentioned earlier (see section 2.1.2),

the effect of the task can affect learners' performance. More reliable results on subject-verb agreement on finite verbs in the present tense could be reached using the same methods and data as in the present study if all the tasks would be included in the inspection. This would also allow comparison not only between different levels of language proficiency but also between different task types.

The complexity of NPs and its effects on the occurrence of subject-verb agreement errors should also be studied in more details. Firstly, there should be more samples of LL in order to be able to show the patterns of errors that they make and to make the results more reliable on a larger scale. Secondly, the experiment should be planned so that it would induce higher numbers of complex NPs which would allow better inspection of the possible reasons for error occurrence. It could then be assessed whether a contrast between the head and the postmodifier and/or the ambiguity of meaning of NP might affect the learners' performance on subject-verb agreement. Learners could also be asked to make grammaticality judgements on sample sentences with correct and incorrect cases. This could reveal learners' actual knowledge on subject-verb agreement better as they might try avoiding difficult structures in their own production, as also Ruin (1996) concluded in her study.

As the present study was cross-sectional, the results showed differences between the levels only at a certain point of time and with a certain type of a task. Compared to the previous studies made on subject-verb agreement, the present study took into consideration also the proficiency level of the learners but the learners were examined as groups according to their language skills. The results then compared the skills of the different groups. To be able to discover actual language development of an individual learner from level to level, a longitudinal study with frequent test points should be done. This way it would be possible to see how the language a learner uses changes, and through what types of changes it goes through to reach higher levels of development. Also the task effect on learner production with individual learners could be

taken into consideration, and as the data collected for the Cefling project included five different tasks so a comparative study could be done already within the same project. These improvements would also show individual variation (see section 2.1.2) within the same proficiency level.

However, the results supported the intuitions that were presented by Wolfe-Quintero et al. (1998: 4, see section 2.3.1) and they also confirmed the hypothesis made earlier. The higher the participants were rated the more advanced texts they produced: they did write more fluently as they wrote more in the same amount of time than the lower level group, they wrote more accurately, and wrote grammatically more complex sentences.

BIBLIOGRAPHY

- Bardovi-Harlig, K. and Bofman, T. 1989. Attainment of syntactic and morphological accuracy by advanced learners. *Studies in Second Language Acquisition* 11:17-34
- Bock, K. and Miller, C. 1991. Broken Agreement. *Cognitive Psychology* [online], 23: 45-93. (14.12.2011)
<http://internal.psychology.illinois.edu/~kbock/bockpubs/Bock%20Miller%201991.pdf> .
- Brown, R. 1973. *A first language. The early stages*. Harvard: Harvard University Press.
- The CEFLING Project (4.2.2012)
<http://www.jyu.fi/hum/laitokset/kielet/cefling/en>.
- The Common European Framework of References for Languages
http://www.coe.int/t/dg4/linguistic/cadre_en.asp (4.2.2012)
- Doughty, C. and Long, M. (eds.) 2003. *The handbook of second language acquisition*. Rowley, MA: Blackwell.
- Downing, A. and Locke, P. 2006. *English grammar. A university course*. Second edition. London: Routledge
- Ellis, R. 1986. *Understanding second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. and Barkhuizen, G. 2005. *Analysing learner language*. Oxford: Oxford University Press.

Eubank, L. 1994. Discussion paper – Towards an explanation for the late acquisition of agreement in L2 English. *Second Language Research* 10: 84-93.

Franceschina, F., Alanen, R., Huhta, A. and Martin, M. 2006. *A progress report on the Cefling project*. Paper presented at SLATE workshop, Amsterdam, December 12, 2011.

Gaies, S. J. 1980. T-unit analysis in second language research: applications, problems and limitations. *TESOL Quarterly* 14: 53-60

Goldschneider, J. and DeKeyser, R. 2005. Explaining the “Natural order of L2 morpheme acquisition” in English. A meta-analysis of multiple determinants. *Language Learning* 51 (1), 1-50.

Housen, A. and Kuiken, F. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30 (4): 461-473.

Hsieh, L., Leonard, L. and Swanson, L. 1999. Some differences between English plural noun inflections and third singular verb inflections in the input: the contributions of frequency, sentence position and duration. *Child Language* 26: 531-543.

Ionin, T. and Wexler, K. 2002. Why is ‘is’ easier than ‘-s’? Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research* [online], 18 (2), 95-136 (13.12.2011)

<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=3&hid=7&sid=a0e4c3a7-a9a1-457c-a82e-c4a6a8202845%40sessionmgr13> .

Larsen-Freeman, D. and Long, M. 1991. *An introduction to second language acquisition research*. London: Longman.

Lightbown, P.M. and Spada, N. 1993. *How languages are learnt*. Oxford: Oxford University Press.

- McLaughlin, B. 1987. *Theories of second-language learning*. London: Edward Arnold.
- Pallotti, G. 2009. CAF: defining, refining and differentiating constructs. *Applied Linguistics* 30 (4): 590-601.
- Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. *Applied Linguistics* 24 (4): 492-518.
- Ruin, I. 1996. *Grammar and the advanced learner*. On learning and teaching a second language. Uppsala: Uppsala University.
- Shibuya, M. and Wakabayashi, S. 2008. Why are L2 learners not always sensitive to subject-verb agreement? *EUROSLA Yearbook* [online] 8 (2) 235. (13.12.2011)
<http://web.ebscohost.com.ezproxy.jyu.fi/ehost/pdfviewer/pdfviewer?sid=a2ac1490-0ba5-4100-ad4c-c1ca5dd29964%40sessionmgr10&vid=2&hid=21> .
- Tarone, E. 1988. *Variation in interlanguage*. London: Edward Arnold.
- VanPatten, B. and Benati, A. G. 2010. *Key terms in second language acquisition*. London: Continuum.
- VanPatten, B. and James, F. L. (eds.) 1990. *Second language acquisition – foreign language learning*. Clevedon: Multilingual Matters.
- Wolfe-Quintero, K., Inagaki, S. and Kim, H.-Y. 1998. *Second language development in writing: measures of fluency, accuracy, and complexity*. Technical report #17. Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.