L1 Use in EFL Oral Proficiency Tests

Lotta Lukkari
Master's Thesis
Master's programme in English studies
Faculty of Arts
University of Helsinki
May 2020

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildningsprogram – Degree Programme	
Humanistinen tiedekunta	Englannin kielen ja kirjallisuuden maisteriohjelma	
Opintosuunta – Studieinriktning – Study Track		
Englantilainen filologia		
Tekijä – Författare – Author		
Lotta Lukkari		
Työn nimi – Arbetets titel – Title		
L1 Use in EFL Oral Profiiency Tests		
Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä– Sidoantal – Number of pages
Pro gradu -tutkielma	toukokuu 2020	40

Tiivistelmä – Referat – Abstract

Tämä pro gradu -tutkielma tutkii koodinvaihtoa suomen kieleen englannin suullisen kielitaidon kokeen aikana. Tutkielma keskittyy koodinvaihdon tarkoituksiin ja määrään. Lisäksi selvitetään, onko oppilaiden suullisella kielitaidolla ja koodinvaihdon käytöllä mahdollista yhteyttä. Aineistona tutkielmaan käytetään HY-Talk- ja FUSE-korpusaineistoja. HY-Talk-aineistoon on kerätty 7.-luokkalaisten ja lukiolaisten puhutun kielen kokeita ja FUSE-aineisto koostuu lukiolaisten puhutun kielen kokeista.

Englannin kielen rooli maailmanlaajuisena lingua francana asettaa uusia haasteita kielenopetukseen. Erilaiset kommunikatiiviset strategiat, mukaan lukien koodinvaihto, ovat hyväksyttäviä jokapäiväisessä kommunikaatiossa, minkä tulisi heijastua myös kielen opetukseen kouluissa. Oppilaiden kohtuullisella äidinkielen käytöllä oppimistehtävien aikana on näytetty olevan edullisia vaikutuksia. Kun koodinvaihto on sallittua sekä kouluympäristössä että sen ulkopuolella, on oletettavaa, että sitä esiintyisi myös koetilanteessa. Suullisen kielitaidon kokeen aikana tapahtuvaa koodinvaihtoa on kuitenkin tutkittu toistaiseksi hyvin vähän, ja tämä tutkielma pyrkii korvaamaan tätä puutetta.

Käytin Antconc-korpusohjelmistoa löytääkseni koodinvaihdot HY-Talk- ja FUSE-korpusaineistoista. Lisäksi selvitin koodinvaihtoa käyttäneiden lukio-opiskelijoiden arvosanat, jotta mahdollista yhteyttä koodinvaihdon tehtävään ja määrään voitaisiin analysoida. Tuloksista selvisi, että oppilaat käyttivät koodinvaihtoa kohtuudella englannin suullisen kielitaidon kokeen aikana. Tulokset osoittavat, että oppilaat käyttivät koodinvaihtoa 12 erilaiseen tehtävään puhutun kielitaidon kokeen aikana. Nämä tehtävät jaettiin kolmeen suurempaan kategoriaan: tehtävän hallinta, sanasto ja muut tehtävän ulkopuoliset toiminnot. Saatujen tuloksien mukaan heikomman kielitaidon oppilaat käyttivät koodinvaihtoa enemmän sanastoon liittyviin toimintoihin kuin korkeamman kielitaidon oppilaat, kun taas korkeamman kielitaidon oppilaat nojasivat koodinvaihtoon tehtävän hallinnallisia tehtäviä varten.

Tutkimuksestani selviää, että Englannin suullisen kielitaidon kokeissa tapahtuva koodinvaihto muistuttaa aiemmissa tutkimuksissa löydettyä, luokkatilanteessa tapahtuvaa koodinvaihtoa. Koska koodinvaihto on hyödyllinen työkalu kommunikaatiolle sekä luokkahuoneessa että sen ulkopuolella, väitän, ettei kohtuullista koodinvaihdosta tulisi rangaista myöskään koetilanteessa ja sitä myöten arvioinnissa.

Avainsanat – Nyckelord – Keywords

Koodinvaihto, englanti vieraana kielenä, korpustutkimus, suullinen kielitaito

Säilytyspaikka – Förvaringställe – Where deposited

Helsingin yliopiston kirjasto

Muita tietoja – Övriga uppgifter – Additional information

Table of Contents

1. Introduction

The English language is in the unique position of a global language with estimated 2 billion people worldwide learning it. It has grown from a national language of less than 3 million speakers to be the global lingua franca in just 500 years (Rose & Galloway, 2019). The use of English is ever on the rise in an increasingly globalized world, which should in turn be reflected in English language teaching (ELT). As native English speakers are the minority of English speakers, the notion of global English teaching becomes more relevant and challenges traditional notions of ELT. For example, Rose and Galloway (2019) note that the students' own language is traditionally seen in ELT as a hindrance and a source of interference, but in what they call global English teaching it is treated as a resource. L1 use should then increasingly be treated as a tool for English language teaching and learning rather than be completely omitted in the classroom.

The VÄISKI project is an undertaking started in the University of Helsinki with the goal of developing the teaching of oral language skills both nationally and internationally. The project focuses on interactivity and pronunciation as the central elements of oral proficiency. The project has conducted, among other activities, workshops and courses for both language teachers and students. A motivation for the project has been the upcoming oral proficiency test to be included in the Finnish matriculation examination, which in turn calls for systematic teaching for oral language skills (Ehrnrooth, 2017). As part of the VÄISKI project, two corpora have been compiled: HY-Talk and FUSE, the latter which is an ongoing and ever-growing project. The two corpora have compiled recordings of Finnish EFL students' oral proficiency examination, HY-Talk of both compulsory basic education and upper secondary school students and FUSE of upper secondary school students. I will be analysing the two corpora and the codeswitching the students use during their oral proficiency examinations.

The use of a student's L1 during an L2 classroom has been shown to have beneficial effects on learning and to be a valued tool for foreign language learning according to students (see e.g. Antón & DiCamilla, 1998; Villamil & De Guerrero, 1996; Neokleous, 2016). Indeed, L1 use is probable when L2 learners share their first language. However, codeswitching to L1 during an L2 oral proficiency test has received little attention. This is understandable considering that in a testing situation only L2 production is expected from the examinees. I argue, however, that if

L1 occurs in the language produced in the L2 classroom, it is only natural for it to occur in a testing situation as well. After all, since codeswitching in a classroom or outside of it is not sanctioned, it is only natural that students would orient towards these language practices in testing situations as well. However, the L1 use happening in L2 English oral proficiency examinations has not really been studied previously. My study then aims to fill this gap. I hypothesise that codeswitching practices in L2 English oral proficiency examination situations resemble the codeswitching practices in EFL classrooms.

In this study, I aim to examine the L1 use of Finnish L2 English students in English oral proficiency examinations. To do this, I have chosen the two Finnish EFL corpora discussed before, HY-talk and FUSE, to use as data for the analysis. I pose the following research questions:

1) What are the functions of L1 use in L2 English oral proficiency examinations?
2) Are there any tendencies of the functions and frequencies of L1 use for students with lower grades versus those with higher grades? If so, what kind?
3) Are there differences in L1 use between HY-Talk and FUSE? If so, what kind?

The first research question focuses on the reasons for which students lean on their L1 during oral proficiency tests. Here I try to find the exact functions of L1 use as well as any tendencies the students may have for the functions used. Answering this question additionally allows me to compare the functions students use their L1 for in classrooms as discovered in previous research to the functions used during oral proficiency tests. Whether they are similar or different and in what ways is one of the main foci of my research. My second research question aims to discover any tendencies of L1 use during oral proficiency tests in relation to the students' oral proficiency as determined by their given grade. I will be examining if a proficiency level affects the frequency of L1 use or the purposes of L1 use. However, no correlation is claimed to be found here, as one cannot be established with this dataset alone. Lastly, my third research question will explore differences between the two corpora, HY-Talk and FUSE. I will focus on how, if at all, the frequency and functions of L1 use differ between the corpora. I will also see how the possible tendencies of L1 use of students of different proficiency levels compare between HY-Talk and FUSE.

The thesis is structured thusly: chapter 2 reviews the related theoretical framework of my study. In the chapter I will explore the concept of codeswitching with special interest in codeswitching by L2 learners. I will also discuss research on L1 use in L2 English classrooms as well as in L2

English oral proficiency examinations more closely. Lastly in the chapter, I discuss language assessment and especially oral language assessment. In chapter 3, I will go into detail on the two corpora, HY-Talk and FUSE, used for this study. In the chapter I also discuss any ethical issues of my data as well as explain my methods for conducting the analysis. Chapter 4 will explore the results of my analysis on L1 use in the HY-Talk and FUSE corpora. The chapter presents the functions of L1 use found from the corpora as well as the found tendencies of L1 use for different language proficiency levels. In the chapter I also discuss the placement of codeswitching in relation to the students' progression in the task. In chapter 5, I will answer my research question by reviewing my results presented in chapter 4. I also return to the literature discussed in chapter 2 and by doing so aim to place my research into the field of research on L1 use by L2 English students. Finally, I conclude in chapter 6 with some closing remarks and the implications of my study.

2. Theoretical Framework

In this chapter, I will review the theoretical framework for the current study. In section 2.1 I will present previous research on codeswitching with specific focus on codeswitching by L2 learners. Furthermore, in section 2.1.1 I focus on literature on codeswitching (to L1) in EFL classrooms during learning tasks, whereas section 2.1.2 examines one study conducted on L1 use during oral proficiency tests. Finally, in section 2.2, I review theory on language assessment and particularly oral proficiency assessment.

2.1 Codeswitching

According to Gardner-Chloros (2009) codeswitching refers to the use of two or more linguistic varieties in the same conversation or utterance by (usually) multilingual people. The phenomenon is widespread and occurs especially in multilingual communities. The term codeswitching was likely coined sometime in the 1950s while the founding body of modern codeswitching research being conducted in the 1960s and 1970s (Benson, 2001). The study of codeswitching has predominantly concentrated on so called true bilingual speakers. That is to say, speakers who have been exposed to two (or more) languages from birth or during early childhood as opposed to through formal education (Bullock & Toribio, 2009, p. 7). However, L2 learners should not be ignored either, a phenomenon related to codeswitching they call style shifting. After all, as Bullock and Toribio (2009) point out, even monolingual speakers are able to switch between linguistic registers and dialects they command. Any person who speaks more than one language to some capacity can codeswitch between them. In my thesis, I will be concentrating specifically on codeswitching to a speaker's L1. This is because my data is consisted of L2 English learners all from Finnish schools and most of them share the same L1 (Finnish). Therefore, codeswitching into the speakers' shared L1 is expected and is the focus of attention in the current thesis.

Relatively little has been written on the codeswitching of L2 learners in general. This is because codeswitching is commonly associated with bilingual speakers who have acquired two or more languages as a child or bilingual communities. However, in principle, the codeswitching tendencies of L2 learners should be reminiscent of those of native speakers. After all, native bilinguals are rarely evenly competent in all their languages similarly to L2 learners. Gardner-

Chloros (2009) notes that many children worldwide are exposed to a second language at school or otherwise take part in some form of multilingual education. Therefore, they are also likely exposed to and contribute to codeswitching happening in a school setting. As the discussion of terminology is common in codeswitching literature, so does Lüdi (2003) discuss whether the L1 used by L2 learners to get around communicative roadblocks is codeswitching, or a phenomenon called translinguistic wording. This idea relates to the majority of codeswitching research being on bilingual speakers and communities, where the languages are acquired as children and not through formal education as is the case with the L2 English learners. It is then brought to question whether the phenomenon of changing from one language to another is the same in both cases. Lüdi notes, however, that even balanced native bilinguals use codeswitching as a strategy to fill in lexical gaps, which are likely to exist no matter the speaker's competence in a given language. For this reason, I will be referring to the L1 use by L2 English learners as codeswitching as well.

My study focuses on codeswitching to L1 in a particular setting: during oral language proficiency examinations. While literature on codeswitching during oral proficiency examinations is scarce, codeswitching in an L2 classroom has been studied quite a bit. It is important to see how codeswitching in classrooms compares to codeswitching in oral proficiency tests, since they are happening in tandem to each other. In the following sections, I will take a closer look at literature on codeswitching (to L1) as it happens in classrooms and in oral proficiency tests in sections 2.1.1 and 2.1.2 respectively.

2.1.1 L1 use in L2 learning tasks

L1 use is likely when two L2 learners share their first language, which is a likely scenario in a school setting. Codeswitching is demonstrated to serve a variety of purposes within the classroom. These include codeswitching as an inclusive strategy for learners of varying language competences or encouraging L2 acquisition to name a few (Martin-Jones, 1995; 2000). Non-excessive L1 use has been shown to have multiple functions and benefits for L2 task completion. Antón and DiCamilla (1998) found L1 use to help maintain interest in the L2 task and to manage the difficulty of the task by developing new strategies. The participants of the study by Villamil and De Guerrero (1996, p. 60) found L1 "an essential tool for making meaning of the text, retrieving language from memory, exploring and expanding content, guiding their action through the task, and maintaining dialogue". L1 has also been shown to

comment on the participants' L2 use, establish joint understanding of the task and formulate the learners' goal (Brooks & Donato, 1994), to scaffold assistance, establish and maintain intersubjectivity (Antón & DiCamilla. 1998), to co-create a shared perspective on the task and to externalise inner speech during difficult activities (Rommetveit, 1985).

Several studies have shown that the benefits of L1 use during L2 tasks depend on the participants' L2 proficiency and the task type. DiCamilla and Antón (2012), for example, found that low proficiency L2 learners use L1 predominantly for task management while high proficiency learners use L1 to discuss vocabulary searches. Task type also affects L1 use. Storch and Wigglesworth (2003) found in their study that L1 was used for task management and task clarification in a joint composition task (meaning focused) while in a text reconstruction (grammar focused) task L1 was employed for meaning or vocabulary deliberations.

Alegría de la Colina and García Mayo (2009) explored the functions of L1 use by low proficiency undergraduate Spanish speaking EFL learners during three different task types: dictogloss (aural stimulus), text reconstruction (written stimulus), and jigsaw (visual stimulus). The students were asked to collaboratively produce a text. In their study, Alegria de la Colina and García Mayo (2009) found two main functions of L1 use: metacognitive talk and metatalk. Metacognitive talk encompassed three main categories:

1) Clarifying or setting task procedures: students discussed the task instructions making sure they had achieved joint understanding.
2) Clarifying and discussing content and meaning: this category included the organisation and sequencing of information or the clarification of information given in a picture or text.
3) Task management: this included comments on labour division, refocusing attention, guiding and monitoring the work, strategy development, and managing affectivity and releasing anxiety (e.g. by using jokes).

(Alegrie de la Colina & García Mayo, 2009, p. 330-332)

Metatalk, on the other hand, included discussions on form. When using metatalk, the students were discussing either vocabulary or grammar. Alegría de la Colina and García Mayo (2009) additionally included the category off-task talk, casual talk unrelated to the task, for comparison purposes. This function was, however, negligible, as there were so few instances of it. This finding suggests that L1 use during EFL tasks does not indicate disengagement from classwork. The study concluded that L1 was employed by the L2 learners as a cognitive tool that enabled

them to access the target language when they lacked the necessary resources while performing tasks.

Storch and Aldosari (2010) investigated the effect on learner proficiency painting and task type on L1 use and the functions of L1 when employed in an EFL learning task. They studied fifteen pairs of Arabic speaking college students who were all asked to complete three tasks. Their study found five functions for which L1 was used:

1) Task management: These included functions of clarifying instructions, recruiting attention, commenting on the quality of the work produced as well as any turns that simply contained a phatic expression.
2) Discussing and generating ideas: L1 was used to generate ideas or to comment on those ideas.
3) Grammar: L1 used to discuss morphosyntax and text structure.
4) Vocabulary: L1 used in deliberations over word or sentence meaning, word searches, and word choice.
5) Mechanics deliberations: L1 used to discuss punctuation, spelling, and pronunciation.

(Storch & Aldosari, 2010, p. 361)

Overall, Storch and Aldosari (2010) found that students used their L1 in moderation when completing the tasks. The most used function of L1 among all the different proficiency pairings combined was task management, which took up 45% of all L1 occurrences. The study found that under task management, the most used function was to negotiate the writing activity and the role of the scribe, though only in pairs of equal proficiency. The second most common function of L1 use were vocabulary items at 26% of all L1 turns. Storch and Aldosari (2010) conclude that L1 use by learners should not be prohibited, as it serves as a valuable cognitive, social, and pedagogical tool.

Azkarai and García Mayo (2015) examined the nature of L1 use and how task modality affects L1 use and its functions. They studied fourty-four EFL Spanish learners across four tasks assigned randomly. They identified five functions for L1 use, which closely corresponded to the functions found by Alegría de la Colina and García Mayo (2009) and Storch and Aldosari (2010):

1) Off-task: Casual talk not related to the task.

2) Metacognitive talk: L1 used for planning, organizing, and monitoring the activity, setting goals, or checking comprehension.
3) Grammar talk: L1 used to discuss grammar.
4) Vocabulary: L1 used to negotiate word or sentence meaning and vocabulary searches.
5) Phatics: Expressions to establish social contact rather than interactional meaning.

(Azkarai & García Mayo, 2015, p. 557-558)

The results showed that overall, the students tended to use L1 in moderation. The most common function was phatics (49.56%) followed by vocabulary (23.23%) and grammar (13.42%). The results of the study differed from those of Alegría de la Colina and García Mayo's (2009), who found that the function of off-task talk in the dictogloss task was minimal, whereas Azkarai and García Mayo (2015) found off-task talk to be the main function of L1 use in the same task type. Additionally, while Storch and Aldosari (2010) found task management functions to be the most prevalent, metacognitive functions were the second least used function in Azkarai and García Mayo (2015).

Overall, previous research has shown that L2 English learners use codeswitching to their L1 in moderation when completing learning tasks in a classroom context. Studies have additionally concluded that L1 use in moderation is a valuable cognitive, pedagogical, and social tool for EFL students and that, therefore, its use should not be completely omitted from the classroom. L2 English learners were found to use L1 for various functions such as task management, vocabulary searches, and off-task functions. L2 English learners have been found to use their L1 for not only cognitive and pedagogical functions but for social functions as well. There is considerable overlap between the different functions L1 is used for found by different studies, which suggests that there are some universalities of L1 use by EFL learners.

2.1.2 L1 use in L2 test interaction

Codeswitching in L2 testing interaction has not been researched much. This is likely due to the expectation of only the target language being produced in oral proficiency test and codeswitching being, therefore, a rare phenomenon in that context. The findings of Storch and Wigglesworth (2003) suggest that students are aware that they should avoid using their L1 in L2 classrooms, and it would be reasonable to assume that the same awareness would be extended to a testing situation. An oral proficiency test, while attempting to simulate natural

spoken language as closely as possible, often is quite a departure from the language used in everyday life. However, codeswitching is a common occurrence in the classroom as a form of task management. I then argue that logically the same language that the students learn and use in the classroom would be the same one used in a spoken test.

Nyroos, Sundholm and Sundqvist (2007) examined codeswitching in a corpus of seventy-nine L2 paired oral proficiency tests for Swedish 9th graders. Their focus was on the initiation of self-repair through using a the Swedish "eller" ("or" in English). Their findings suggest that "eller"-initiated repair serve to display trouble awareness, amplify the test-takers' attention to what needs to be revised or replaced, indicates to co-participants that self-repair is underway, push forward turn transition, and pre-empt conclusions about the speaker's stance or linguistic competence, the latter of which is particularly relevant in a language testing context. Nyroos et al. (2007) then argue that the codeswitching of single particles is not, against a common folk belief, accidental and something done under pressure, but rather a form of meta-commentary or even a brief moment of self-directed talk. It is, however, unclear how Nyroos et al. (2007) reach the conclusion of the codeswitching of single particles is not accidental. The presented instances of "eller" are not always marked e.g. with hesitation. When hesitation is present, it is not clear whether it marks the codeswitching to Swedish or other difficulties in producing the target language. Pietikäinen (2014) discussed the concept of automatic codeswitching in her study on ELF couples' codeswitching tendencies. She defines automatic codeswitching as instances of codeswitching not explicitly marked by e.g. hedging or other such flags and happen passively (p. 13). I would then argue that the instances of the Swedish "eller" in Nyroos et al. (2007) are, in fact, automatic or accidental. I will return to the results of Nyroos et al. (2007) again in chapter 5, in which I will also compare my results to theirs.

2.2 Spoken Language Testing and Assessment

Speaking skills are an important part of the Finnish national curriculum of language teaching and therefore a salient object of assessment as well. Luoma (2004) describes the development of speaking tests as a cyclical process: First, spoken assessment tests are deployed to fulfil a need for scores for a purpose. Next, a test is developed, administered, and scored. The resulting speech, test discourse, is then assessed in relation to the assessors and the criteria used, and a score is produced. The cycle is then begun anew. Spoken assessment can, however, be quite challenging due to the plethora of factors influencing impressions of speaker competence.

According to Luoma (2004), one of these factors is the task type used, which has an effect on the test discourse. For instance, Luoma (2004) discusses differences of two spoken task setups: an examiner acting as an interviewer for a test-taker and a pair of test-takers speaking with each other. An advantage of a paired spoken test over an interview test is that it eliminates the unnatural interviewer-interviewee roles and allocates the speaker responsibilities more evenly across both speakers. In her study, Brooks (2009) found that test-takers obtained better scores and produced more complex test discourse in the paired format with another test-taker when compared to the individual format being interviewed by an examiner. Paired work has also been shown to encompass more interactional functions, such as turn-taking and topic management, and interactive listening (Ducasse & Brown, 2009). Paired work can have its disadvantages, though, as Luoma (2004) points out. Namely, the examinees performances are inevitably affected by the other participant's personality, communication style and possibly even language competence. This can work to both make the speaker seem more or less competent than what they actually are.

Luoma (2004) discusses the challenges of reliable assessment of speaking. The interactive nature of speaking makes the resulting test discourse unpredictable. The same unpredictability also concerns the human assessor and the subsequent rating process. Luoma (2004) describes the procedures taken to combat this resultant unreliability for classroom assessment specifically, in which subjectivity is a common worry. Subjectivity can be reduced by assessing performance anonymously one task at a time. However, this becomes a problem when assessing spoken tasks, since the task by task rating is impractical and anonymity is negated by each test takers recognisable voice even from just an audiotape. Hence Luoma (2004) infers that the only alternative is for the assessor to critically reflect upon their own rating work.

There is not much literature on how codeswitching to L1 in an oral proficiency test affects assessment if at all. Assessors of the HY-Talk project noted a Finnish coded hesitation ("öö") as a negative factor in terms of fluency, and sounding Finnish was also seen as a detriment (Itkonen, 2010, p. 52). The possible implications of codeswitching on the final grading process are not explicitly discussed in Itkonen's thesis, however.

As has been illustrated, language assessment, and especially that of spoken language, can be a tricky feat. This is due to the multitude of factors on every level of assessment including test development, the test-takers, and the individual assessors. It is unclear how, if at all, a test-taker's codeswitching to their L1 during oral proficiency examination affects assessment of oral

production. As Bullock and Toribio (2009, p. 1) note, codeswitching is commonly associated by the general public as indexing language degeneration. While linguists view codeswitching as a marker of bilingual proficiency, the common belief of codeswitching as a signal of gaps in language proficiency may still influence oral language assessment negatively. However, without proper empirical research, the effects of codeswitching on language assessment will remain unknown. In order to study a possible correlation between codeswitching and language assessment, we must first understand the codeswitching used by L2 students during oral proficiency examinations. This is the goal of the current study.

3. Methodology

In this chapter, I will introduce the material used as data in the current thesis as well as the methods used for analysis. In order to study codeswitching in the Finnish L2 oral proficiency test situations, two Finnish EFL corpora were selected for analysis: Hy-talk and FUSE. I will begin by introducing the HY-talk corpus in section 3.1, followed by the FUSE corpus in section 3.2. In section 3.3 I discuss limitations of my dataset and ethical issues in section 3.4. Lastly, in section 3.5, I will describe the methods which I used to conduct the analysis.

3.1 HY-talk Corpus

The HY-talk corpus was a 3-year research project carried out by the University of Helsinki on oral proficiency in compulsory basic education and general upper secondary education. The project was conducted in order to investigate and improve reliability of the oral proficiency scales included in the Finnish core curriculum used as a tool for assessment of communicative oral proficiency. The database includes transcriptions of examinees taking the oral proficiency test in English, Swedish, German, and French. However, for the purposes of the current thesis only the English transcriptions will be referred to. The 53,115-word English section of the HY-talk corpus consists of transcriptions of oral proficiency tests taken by eight lower secondary school and twenty upper secondary school pairs from three different schools in the Helsinki metropolitan area in 2007. The tests were also recorder in both audio and video format by HY-talk employees conducting the test. For this study, however, these formats will be ignored as only the transcriptions are necessary to analyse since any occurring codeswitches are tagged with non-embedded tags.

The tasks used in the oral proficiency test were designed specifically to be used for the HY-talk project. Both lower secondary school and upper secondary school tests were structured similarly, though the subject matter is different. Before the recording, the examinees were given twenty minutes to familiarise themselves with the tasks. The test comprised of four set of tasks: a warm-up interview, a monologue task, a set of three structured dialogue tasks, and an open-ended dialogue task as an extra. The structured dialogue tasks included one presentation and three imaginary communicative situation which the examinees were required to act out. For the

extra task, the examinees were asked to have a more open discussion on the topic relating to reality TV. This task was optional, but most of the participants chose to do it.

The performances of the examinees were assessed by several language experts. The students were given a grade based on five skill areas: (1) performance, (2) fluency, (3) pronunciation, (4) range, and (5) accuracy. Additionally, an overall grade was given to the examinees based on their performances. The grades were documented in a separate file with the designated student identity numbers corresponding to the ones in the transcriptions' metadata.

3.2 FUSE Corpus

FUSE (The Finnish Upper Secondary School Corpus of Spoken English) is an on-going web-based open-access corpus project comprised of transcribed recordings of Finnish upper secondary school students taking the oral examination test at the end of the vocational course 8 (Ehrnrooth, 2018). A total of twenty-eight conversation have been recorded in the FUSE corpus as of the time of writing. FUSE invites Finnish EFL teachers to participate in its compilation by submitting recordings of their own students through the FUSE website. The earliest recording in the corpus is from 2014, with new entries still being added. Due to the project being ongoing, the corpus is still relatively small with only 20,392 words. Similarly to the HY-Talk corpus, instances of codeswitching have been tagged in the transcription with non-embedded tags.

The transcriptions in the FUSE corpus represent the third task in the Spoken Examination of English for Finnish upper secondary school students. The subject matters of the tasks vary, but there are only two task structures: structured dialogue and mind-map supported conversation. The corpus includes assessment of the examinees' performances in the third task of the test by the teacher conducting the test as well as university students in the University of Helsinki MA programme in English Studies for some examinees. For the sake of clarity, I will only consider the teachers' assessment of the students in my analysis. The participants were graded based on 5 skill areas: (1) interaction, (2) fluency, (3) pronunciation, intonation and stress, (4) vocabulary and structures, and (5) grammar. From these, an average score was also given. The teachers' grading is given in both school grades and based on the scales in the Common European Framework of Reference for Languages (CEFR).

3.3 Limitations of the Data

The dataset poses some limitations on my analysis, which should be taken into account. My analysis concentrates on the transcriptions of the recordings done for both HY-Talk and FUSE. Although I have access to the audio recordings in FUSE and at times could benefit from them when determining the function of L1 use in a conversation, with HY-Talk I was limited to the transcribed version as I did not use the audio or video recordings to any extent. This may limit my understanding of the conversations and specifically the functions of L1 use, as I am missing some aspects of the discourse such as any non-linguistic elements like facial expressions and body language. It should also be noted that due to the nature of corpora, I am limited to the information presented to me. I have no knowledge of what happened before or after the recordings, the former which becomes quite relevant in section 4 of my thesis.

Another limiting element of particularly the HY-Talk corpus is the inconsistent tagging of some codeswitches. This was particularly a problem in the case of proper nouns (further discussed in section 4.1), as sometimes an instance was tagged as L1 use and at others not. It seems that the majority of L1 instances were tagged, but in the case a codeswitch was left with no tags, it has not been detected by the corpus software and therefore was left out of my analysis.

3.4 Ethical issues

In order to protect the privacy of the students who took part in both HY-Talk and FUSE, the data has been pseudonymised. The students in both corpora are appointed their individual codes with which to distinguish them from each other. Furthermore, HY-Talk is not available publicly and can only be accessed as a local file. Any copies of the data are destroyed after analysis is completed.

3.5 Methods

In order to analyse the occurrences of codeswitching in the HY-talk and FUSE corpora, the freeware corpus software AntConc (Anthony, 2019) was used. The transcriptions of both corpora were transferred into text file format (.txt) with all metadata omitted to ensure reliable results in the AntConc software. The searches for instances for codeswitching were done separately for each corpus due to different tagging systems. The results were then exported into

two separate spreadsheet files (.xlsx) for the Microsoft Excel. In order to deduce what functions the codeswitches to L1 served, the found instances of codeswitching were manually tagged according to their perceived functions. At this point the reference for these functions were based purely on the data while referencing any previous research mainly when deciding what to call a specific function I have omitted instances of codeswitching that were by the instructors, marked as inaudible, or part of a previous instance but cut off in the transcription (e.g. due to overlapping speech). The instances of codeswitching were additionally categorized according to where they appeared in the tasks: before or at the beginning, during, or at the end of the task. This was done so that comparison could be drawn between the two corpora and the tendencies of L1 use. Lastly, the examinees were also tagged by their individual student IDs as well as their overall CEFR-based scores. The Microsoft Excel's pivot-table functionality was used to count the different instances to see how the functions of L1 are distributed through the corpora. The normalized frequencies of the occurrences of codeswitching were counted in order to draw comparison between the two corpora.

4. Results

In this chapter, I will present the results of my analysis of the transcribed oral proficiency tests in the HY-Talk and FUSE corpora. I will begin with explanations of the functions found in both HY-Talk and FUSE. In section 4.1 I will introduce the functions of L1 use I found in both HY-Talk and FUSE and my categorisation of them. In section 4.2 I will go over the functions of L1 use found in the HY-Talk corpus, followed by section 4.3 in which the functions of L1 use in FUSE are explored. In section 4.4, I will compare the L1 usage and the students' grades in order to find any possible correlations. Finally, in section 4.5, I will analyse the instances of codeswitching in terms of at what point of the task they appear, whether it is in the beginning, middle, or end of a task.

4.1 The Functions of L1 use

I found a total of twelve different functions for which L1 was used in the HY-Talk and FUSE corpora. These functions come purely from the dataset and are not necessarily based on any previous research on codeswitching. This was done in order to not be needlessly influence the results I get from my data. I am inclined to also mention that the following categorisation is relatively subjective as it is based solely on my own understanding of the data and the context of the instances, which is in turn limited to the written transcriptions of the recorded conversations. I have, however, gone through the data several times both with the corpus-software as well as through close reading reviewing my categorisation multiple times. I therefore claim my categorisation to be reliable.

The functions found in the dataset can be divided into three main categories: task management, vocabular, and other. These categories are further divided into sub-categories, which are introduced with examples in this section.

Task management functions work to in some way enable the completion of a task. A requisite for an instance of L1 use to be categorised as task management was that it had to somehow directly relate to the task at hand. The category is divided into four sub-categories, which are as follows:

1) **Checking comprehension** occurs when students check whether they have understood either something in the task or something their co-participant said.

(1) Siis eiks ne niinkö tienny et ne oli siskoja? [So did they not know they were sisters?]

2) **Clarifying instructions** refers to instances when students ask for clarification on what the task description means or what they are required to do. Requests of clarification are usually directed at the instructor but can also be for the co-participant.

(2) *Hei mitä mun pitää tehä tässä*? [Hey what do I need to do here?]

3) **Organizing activity** refers to the students codeswitching in order to move forward in the task, such as initiating the task or prompting the co-participant to move onto the next section.

(3) Alotetaaks me vaan? [Do we just start?]

4) **Turn allocations** are codeswitching used specifically to decide who's turn is to speak or who plays what role. While this could be allotted under organizing activity, turn allocation was such a common function, especially in the HY-Talk corpus, that a separate category was created.

(4) *Eiku alota sä.* [No you start.]

Vocabulary functions focus on the language itself and more specifically on individual words. The items under this category were used to either signal trouble awareness and preface self-correction or to request assistance. The category is further divided into two sub-categories:

1) **Correction** refers to instances where a student signals awareness of an error, which is usually followed by self-correction. A typical example of correction is the Finnish "eiku" (roughly the Finnish equivalent of the English "I mean").

(5) My father name is […] she *eiku* [I mean] he is er 42 years old.

2) **Word retrievals** are instances of codeswitching where the students cannot think of a word in the target language. Instead, they produce the desired word in Finnish or explicitly ask what a word is in English. The students typically either produce the word

in the target language shortly after or receive help from their co-participant or the instructor. In some instances, however, the students do not produce the word in the target language but mark the utterance with e.g. hesitation or laughter implying the they are aware of their L1 use. This distinguishes word retrievals from word slips, which are discussed later in this section.

(6) The- it's er *mikä on linnoitus*? *[What is a fort?]*

The **other** category consists of functions that either do not fall under the other two main categories or are otherwise unrelated to the task. This category was devised for instances of codeswitching that were clearly separate of the English utterances or that would otherwise needlessly skew the numbers of some other category distorting my results. The category is divided into six sub-categories.

1) **Off-task talk** refers to speech that is not directly related to the task at hand. Typical off-task talk in the data is expressing opinion of the task, most likely to alleviate anxiety or otherwise to express frustration. The function of releasing anxiety was categorised as task management under metacognitive talk by Alegría De La Colina and Del Pilar García Mayo (2009). They also categorised off-task talk as a separate entity. However, because this function does not directly contribute to the completion of a task, I have chosen to include this function under the category of other.

(7) Tää on hirveetä. [This is awful.]

2) **Phatic expression** are expressions that do not hold meaning on their own. The instances of codeswitching that do not seem to serve any interactional function were categorized as phatic expressions. They are also separate from the English utterances. Storch and Aldosari (2010) have phatic expressions under task management. However, just like off-task talk, I do not consider phatic expressions to be directly contributing to task completion. Therefore, I have decided to categorise it here.

(8) It would be *joo* [yeah] however what about the last…

3) **Proper nouns** are nouns referring to an entity, such as an amusement park or a TV show. Some of the proper nouns used have an English counterpart and the students do not know it, while others do not (e.g. “Linnanmäki”) but were nonetheless tagged as codeswitching.

(9) And erm then in the centre and, then the erm *Tuomiokirkko* [Helsinki Cathedral].

4) **Responses** in this data set are Finnish utterances used to respond to an earlier statement, which typically is also in Finnish. Since responding to a Finnish statement in Finnish is expected, this category was created in order to not needlessly bloat other categories. Interestingly, this category includes instances of students sometimes answering the instructor in Finnish even though the instructor spoke English.

(10) S1: I think they they were very er *mikä se sana oli* [what was that word]?

S2: Mikä? [What?]

5) **Signalling trouble** refers to instances where students express that they are having trouble completing their task for one reason or another. However, these instances are not explicitly asking for assistance of clarification of the task and therefore warrant their own category separate from e.g. *word retrieval* or *clarifying instructions*.

(11) Ah *mä en muista tästä yht-* [I do not remember any of this] er oh my god it's a lovely room…

6) **Word slip** are uses of L1 as a part of an English sentence characterised by the students not seeming to notice their codeswitch as many of the instances are not marked by e.g. hesitation or laughter. In some cases, the slips are caught by the students and they correct themselves with the corresponding English word and in other cases the Finnish word slips by unnoticed. Typical words in this category are conjunctions (e.g. “ja”, Finnish “and”) or so-called filler words (e.g. “no”, Finnish “well”).

(12) I can’t really say who is my best friend *koska* [because] ‘cause i have so many friends that I know.

In the following sections, I will discuss these functions as they are used in the HY-Talk and FUSE corpora.

4.2 Functions of L1 use in HY-Talk

Starting with the HY-Talk corpus, out of the twenty-eight conversations, twenty-two featured codeswitching to L1. 227 separate instances of L1 use were found with the normalized frequency of 42,7 per 10 000 words. All twelve functions presented in section 4.1 were found in the HY-Talk corpus. The number of the different functions can be seen in table 1:

Functions of L1	**Amount of Instances**
Turn allocation	53
Word retrieval	37
Clarifying instructions	32
Phatic expression	28
Word slip	15
Response	15
Off-task talk	10
Signalling trouble	10
Organizing activity	8
Checking comprehension	7
Proper noun	7
Correction	5
Total	**227**

Table 1. Functions of L1 use in the HY-Talk corpus.

Most of the functions fell under the function of turn allocation. Other major categories included word retrieval, phatic expressions and clarifying instructions. The L1 functions used the least are comprehension checks, corrections, and proper nouns. The latter is, however, of least interest. This is due to the inconsistent tagging in the corpus with some proper nouns being marked as codeswitching and others not. Moreover, proper nouns being names that often do not have an English counterpart hardly even count as codeswitching. However, since they were tagged in the corpus they are also displayed here for maximum recall. The distribution of the different functions with percentages can be viewed in Figure 1.

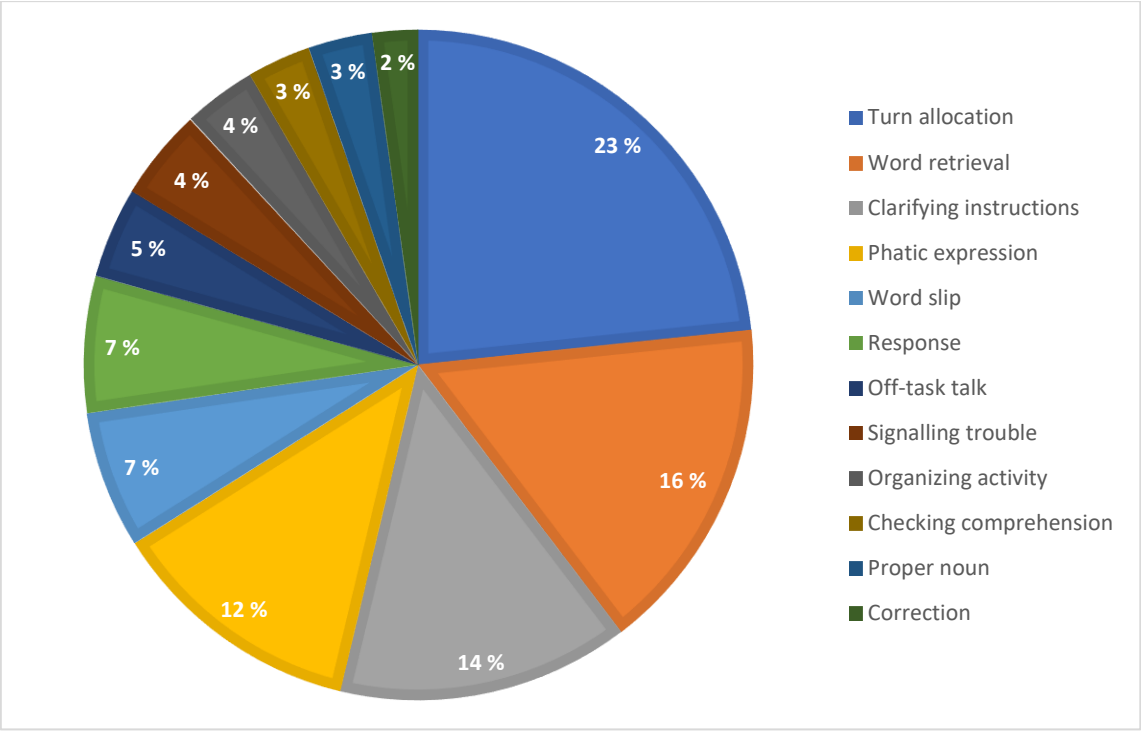

Figure 1. Distribution of the functions of L1 use in the HY-Talk corpus.

Overall, 40% of the functions used in HY-Talk fall under task management (checking comprehension, clarifying instructions, and turn allocation), while 16% belong under vocabulary functions (correction and word retrieval). The remaining 42% of the instances belong under the other section (word slip, response, off-task talk, phatic expression, signalling trouble, and proper nouns).

Of the functions under task management, by far the most prominent was turn allocation, which took up almost half of the category. As discussed in section 4.1, the sheer number of turn allocations prompted me to create a separate sub-category for them as otherwise they would simply fall under organizing activity. It is impossible to say for certain which circumstances led to the students discussing turn taking during the task instead of before the recording started, since my analysis is limited to the information on the transcriptions.

4.3 Functions of L1 use in FUSE

From the twenty-eight conversations collected so far for the FUSE corpus, only nine conversations included L1 use. From these nine transcriptions, a total of twenty-six separate instances of L1 was found. These numbers are considerably less than what was found in the

HY-Talk corpus, in which twenty-two out of the twenty-eight conversation included a total of 227 instances of codeswitching. The normalized frequency of the L1 instances is 12.8 per 10 000 words, making codeswitching a much less common occurrence than in the HY-Talk corpus. In addition to the amount of L1 use being lower than in HY-Talk, there are also far less variation of different functions in the FUSE corpus. Whereas 12 functions were found in HY-Talk, FUSE exemplified only 7, almost half the amount of functions in HY-Talk. There were no additional functions found in FUSE, as the 7 functions found in FUSE were also found in HY-Talk. This indicates that L1 is employed in a similar fashion by all (Finnish) EFL students. The functions and their amounts found in FUSE are presented in table 2.

Functions of L1	**Amount of Instances**
Word slip	10
Phatic expression	6
Word retrieval	4
Organizing activity	2
Response	2
Signalling trouble	1
Proper noun	1
Total	**26**

Table 2. The functions of L1 use in the FUSE corpus.

Furthermore, the percentual distribution of the functions of L1 use in FUSE are illustrated in figure 2.

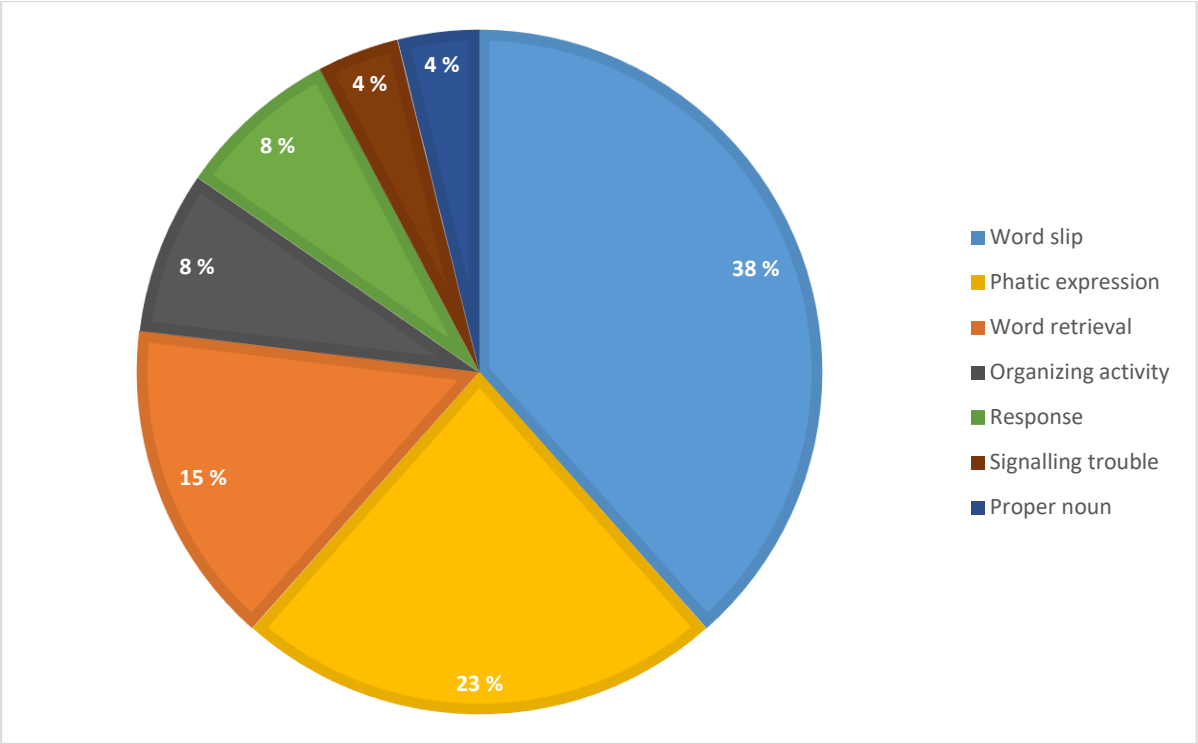

Figure 2. The distribution of functions of L1 use in the FUSE corpus.

The most prominent function for which L1 is used in FUSE seem to be word slips followed by phatic expressions, which are both characterized by their apparent accidental nature. That is to say, both these categories feature instances of L1 use not explicitly flagged and therefore appearing to occur passively. Together they make up 61% of all instances of L1 use in FUSE. This indicates that a major portion of codeswitching happening in the oral proficiency test context in FUSE is accidental and perhaps even done unconsciously without the speaker ever noticing. This is not to say all instances of L1 in FUSE use are involuntary. Word retrieval, organizing activity, and signalling trouble are all apparently voluntary functions and take up 27% of all instances.

In terms of the three broader categories, student in the FUSE corpus used other, off-task functions the most. The category of other took up 77% of all functions. The second most used category was vocabulary at 15% and lastly the least used category was task management at 8%. This indicates that the students in FUSE leaned on their L1 the least for functions directly having to do with the task at hand or the English language itself. Rather, they used Finnish mostly for building social relations. While the category of other was also prominently featured in the results of the analysis in HY-Talk, it is interesting how the category of task management compares between the two corpora. Whereas task management was quite rare in FUSE, in HY-Talk the category took up 44% of all L1 turns.

As discussed before, five of the total twelve found functions are absent in FUSE: turn allocation, clarifying instructions, off-task talk, checking comprehension, and correction. This is especially interesting since turn allocation and clarifying instructions were so prominently featured in HY-Talk. When discussing the placement of the codeswitches in a task in section 4.5, I will further explore this difference and possible reasons for it.

4.4 L1 Use and Grading

In this section I will explore any possible correlation between the CEFR-based grading of the upper secondary school students of HY-Talk and FUSE and the functions and frequencies of codeswitching. For this analysis, only the upper secondary school students' results are analysed. This was done for a couple of reasons. First, the HY-Talk transcriptions for compulsory basic education students lacked the individual student IDs, which prohibited the connection of grade and student. The student IDs were given for the upper secondary school student. Second, since FUSE consists only of upper secondary school students, choosing only the upper secondary school grades of HY-Talk allows for meaningful comparison. It should also be noted that it is impossible to definitively establish a correlation between codeswitching and grading in this case. Grading is a complex process which is affected by multiple elements of both the students' performance as well as the individual grader. Whether codeswitching affected the grading process or not is unclear. The aim of this analysis is then to see whether a student's proficiency level (as determined by the grade given to them) has any bearing on their L1 use tendencies. However, no absolute claims of correlation are made.

There was a total of 23 upper secondary school students in HY-Talk who used L1 during their oral proficiency test. Two of them were graded at A2.2, thirteen at B1.1, four at B1.2, three at B2.1, and one at B2.2. Figure 3 illustrates how the amount of codeswitches as well as the different functions of them are divided between differently graded students in the HY-Talk corpus.

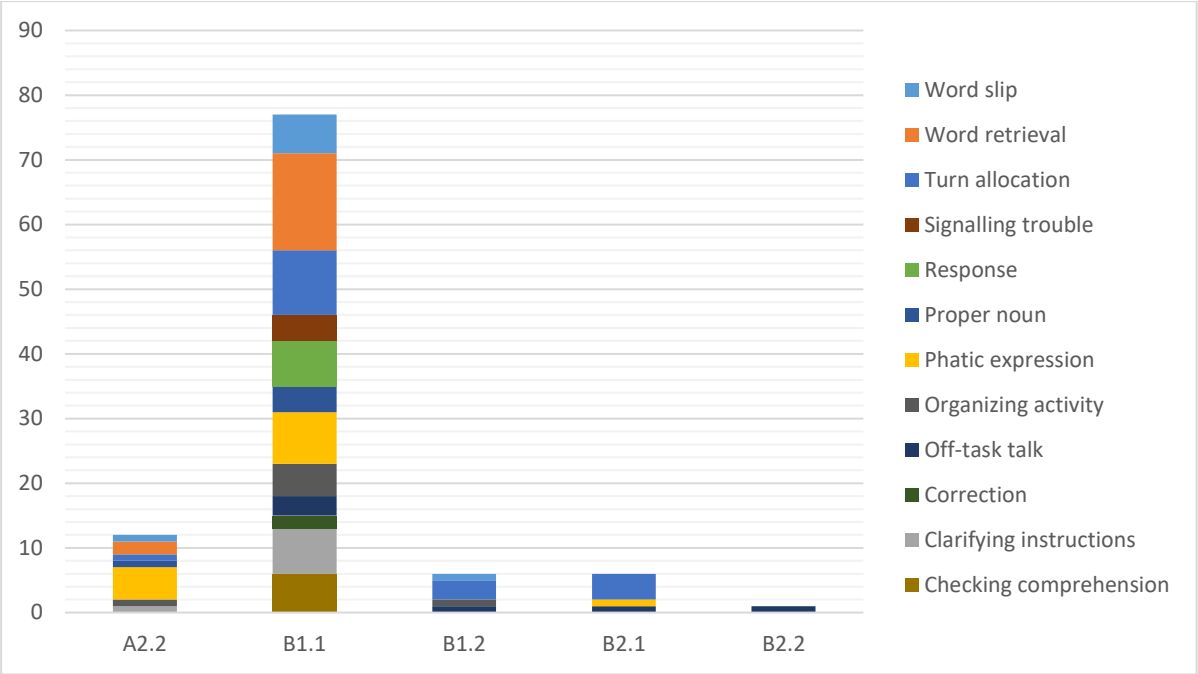

Figure 3. Amount of codeswitches across different grades in HY-Talk.

The vast majority of L1 use comes from students graded with B1.1. However, this large gap is due to most of the students being graded thusly (13 out of 23) and additionally a couple of individuals in this group contributing a large number of codeswitching into the dataset. Due to the uneven division of grades normalized frequencies were counted. L1 was used most frequently by students graded with B1.1 (0.70 L1 turns per 100 words), closely followed by A2.2 (0.69 L1 turns per 100 words). At higher grades, the frequency of L1 use lowered quite drastically. B2.1 graded students (0.19 L1 turns per 100 words) used codeswitching slightly more than students graded B1.2 (0.12 L1 turns per 100 words). L1 was used the least by students with the highest grade, B2.2 (0.08 L1 turns per 100 words). These findings suggest that lower proficiency does not automatically mean more codeswitching, since the students with the second lowest grades took the most L1 turns. It is clear in these results, however, that the amount of codeswitching lessens the higher the student's oral proficiency

Table 3 presents the percentual distribution of different functions for which L1 is used for. The findings show most variety of functions used at grade B1.1, with each found category being used. Although, this result could also be contributed by the large number of speakers in the group.

	A2.2	B1.1	B1.2	B2.1	B2.2
Task management	**24.99%**	**36.36%**	**66.67%**	**66.67%**	**0.00%**
Checking comprehension	0.00%	7.79%	0.00%	0.00%	0.00%
Clarifying instructions	8.33%	9.09%	0.00%	0.00%	0.00%
Organizing activity	8.33%	6.49%	16.67%	0.00%	0.00%
Turn allocation	8.33%	12.99%	50.00%	66.67%	0.00%
Vocabulary	**16.67%**	**22.08%**	**0.00%**	**0.00%**	**0.00%**
Correction	0.00%	2.60%	0.00%	000%	0.00%
Word retrieval	16.67%	19.48%	0.00%	0.00%	0.00%
Other	**58.34%**	**41.56%**	**33.34%**	**33.33%**	**100.0%**
Off-task talk	0.00%	3.90%	16.67%	16.67%	100.00%
Phatic expression	41.67%	10.39%	0.00%	16.67%	0.00%
Proper noun	8.33%	5.19%	0.00%	0.00%	0.00%
Response	0.00%	9.09%	0.00%	0.00%	0.00%
Signalling trouble	0.00%	5.19%	0.00%	0.00%	0.00%
Word slip	8.33%	7.79%	16.67%	0.00%	0.00%

Table 1. Percentual distribution of functions of L1 use across different grades in HY-Talk.

According to the findings it would also seem that lower proficiency learners employ LI for vocabulary searches more than higher proficiency learners, while learners of higher proficiencies rely on their L1 mainly for task management purposes. The category of other is also prominently featured at all proficiency levels. The results from the grade B2.2 consist of only one speaker with only one L1 turn and cannot therefore be deemed reliable.

In FUSE, there was a total of 12 speakers who used codeswitching of which three were graded at B1.2, eight at B2.1, and one at B2.2. The variation of grades given in FUSE is more limited than in HY-Talk. This allows for less comparison between different proficiencies and further prevents any conclusive findings on the relationship of oral proficiency and L1 use. Regardless, some tendencies seem to emerge from my findings. Figure 4 shows the amount of codeswitching as it divides between the different oral proficiencies. However, in order to compare the different proficiencies and their L1 usage, normalized frequencies of codeswitching were counted. Students graded B1.2 used codeswitching the most according to the results with 0.94 L1 turns per 100 words. This number is quite large especially when compared to the results from HY-Talk. It should be noted, however, that the word count in FUSE is considerably lower than in HY-Talk which likely skews the results somewhat. The differences can also be explained in part by individual differences of the individual students.

The second most frequent user was the highest grade, B2.2, with 0.52 L1 turns per 100 words. This result, however, cannot be deemed reliable, because the group consists of a single speaker whose wordcount was relatively low (193 words). Additionally, this result comes from a single L1 turn. Lastly, the students graded B2.1 took 0.19 L1 turns per 100 words.

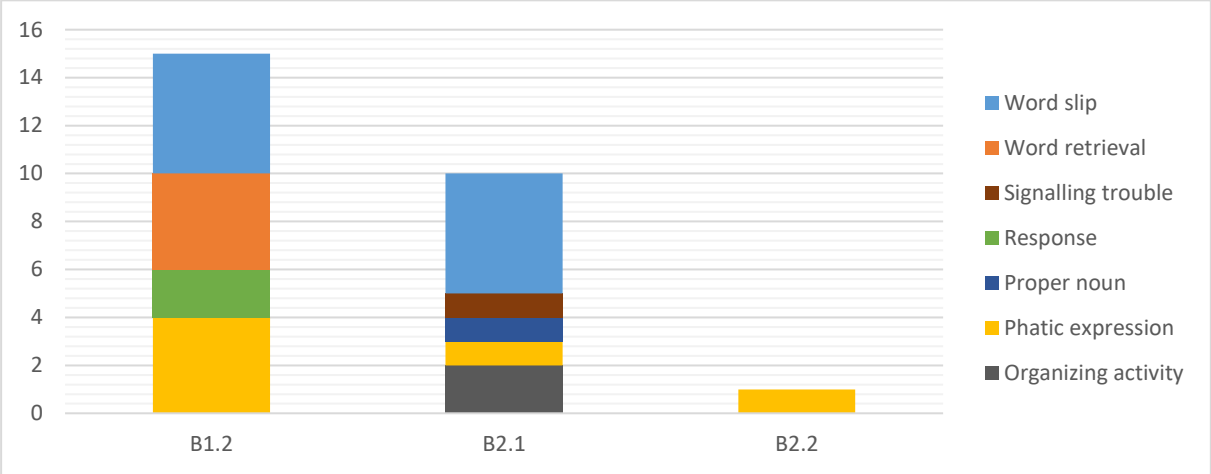

Figure 4. Amount of codeswitches across different grades in FUSE.

Next, I examined the functions for which L1 was used by students of different proficiency levels. The percentual division of the functions are presented in table 4. Based on these results, it appears that vocabulary functions are used more by lower proficiency learners than those of higher proficiencies. Task management was most common for learners with the grade B2.1. The most common functions for which L1 was used for across all proficiencies fall under the category of other. However, the small size of FUSE makes it difficult to draw any definitive claims based on these results.

	B1.2	B2.1	B2.2
Task management	**0.00%**	**20.00%**	**0.00%**
Organizing activity	0.00%	20.00%	0.00%
Vocabulary	**26.67%**	**0.00%**	**0.00%**
Word retrieval	26.67%	0.00%	0.00%
Other	**73.33%**	**80.00%**	**100.00%**
Phatic expression	26.67%	10.00%	100.00%
Proper noun	0.00%	10.00%	0.00%
Response	13.33%	0.00%	0.00%
Signalling trouble	0.00%	10.00%	0.00%
Word slip	33.33%	50.00%	0.00%

Table 2. Percentual distribution of functions of L1 use across different grades in FUSE.

From the results from the usage of L1 by students of different oral proficiencies in both HY-Talk and FUSE, it seems that vocabulary functions are favoured by learners of lower proficiencies, while the L1 use of higher proficiency learners leans more towards task management. Other functions are common across all proficiency levels in both corpora, so much so that these functions are the most common in FUSE overall.

The result from both corpora have the disadvantage that the group with the highest grade only has one student as well as only one instance of L1 use, which makes meaningful analysis hard. The small number of speakers in FUSE especially also makes it difficult to make definitive claims on the possible relationship of spoken language proficiency, grading, and codeswitching. It is possible that individual differences here weigh more than differences in language proficiency. Further research on the subject is needed.

4.5 The Position of L1 Use During Task Completion

In this section, I will go over the functions of codeswitching in terms of when they occur during completing a task. I have categorised the functions to either occurring in the beginning, middle, or end of a task. An instance categorised to occur in the beginning happened either before beginning the task or right after starting (e.g. after the first sentence). Instances occurring in the middle are uttered well into the task completion. The instances occurring at the end mark an ending of the task or the ending of the speaker's last turn during that task.

The positions of the instances of codeswitching in the HY-Talk corpus are illustrated in figure 5.

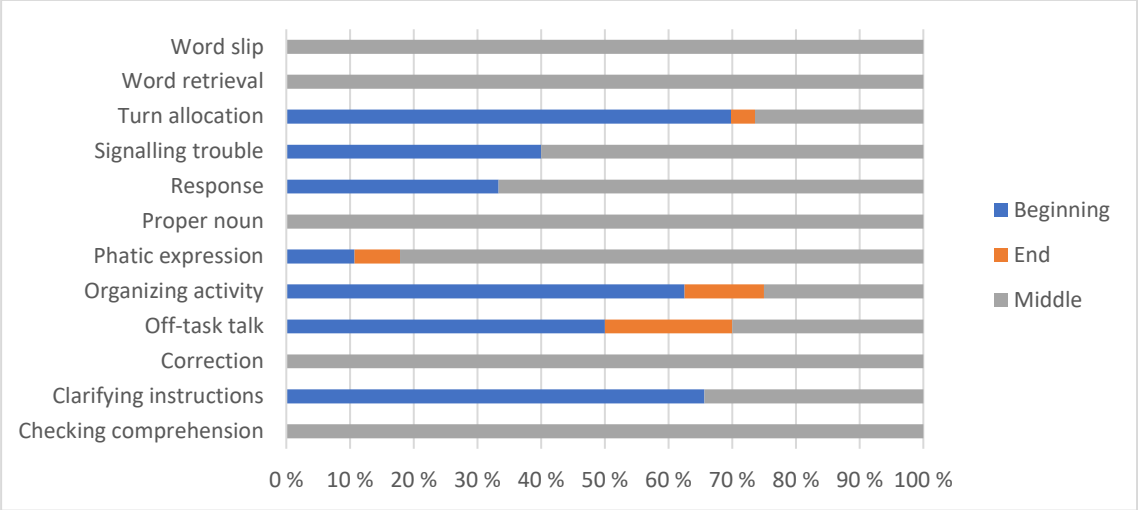

Figure 5. Positions of codeswitches within a task in HY-Talk.

From this we can see that turn allocation, clarifying instructions and organizing activity all have a tendency to occur either before the task or right at the beginning of it. All of these functions fall under task management. In HY-Talk it was common for the students to discuss the tasks before starting either with each other or with the instructor. Surprisingly, comprehension checks only occurred in the middle of the tasks, even when it was regarding the task description. Functions having to do with vocabulary and grammar seem to occur only during the completion of a task. These functions include word slips, word retrievals, and corrections. Students would occasionally mark the end of their task with codeswitching. However, there is no consistent function used to end a task. In the HY-Talk data, four different functions were employed to signal the end of a task: turn allocation (e.g. “sun vuoro”, Engl. “your turn”), phatic expression (e.g. “joo”, Engl. “yes”), organizing activity (e.g. “emmätiiä kai se oli siinä sitte”, Engl. ”I don’t know I guess that was it”), and off-task talk (e.g. “voi juma hirveetä”, Engl. “oh gosh awful”).

Since the task management functions of L1 use in HY-Talk typically occur before or right at the beginning of a task, an argument could perhaps be made of them not being codeswitching during an oral proficiency test at all. I, however, do argue that they can be perceived to be a part of the test situation. Although precise circumstances are unknown to us, we know that the students had at least some time to prepare before the recordings. We also know that the decision to start the recording was made and still task management in the students’ L1 occurred. The decision was also made to include these occurrences in the transcriptions. Therefore, I also include them in my analysis as well. It is still valuable to know when the codeswitching occurs

within the timespan of task completion, as some functions seem to tend to occur at one time over another. This will also come apparent when we move on to examine the placement of L1 use found in FUSE next.

The positions of codeswitching occurring in the FUSE corpus are illustrated in figure 6:

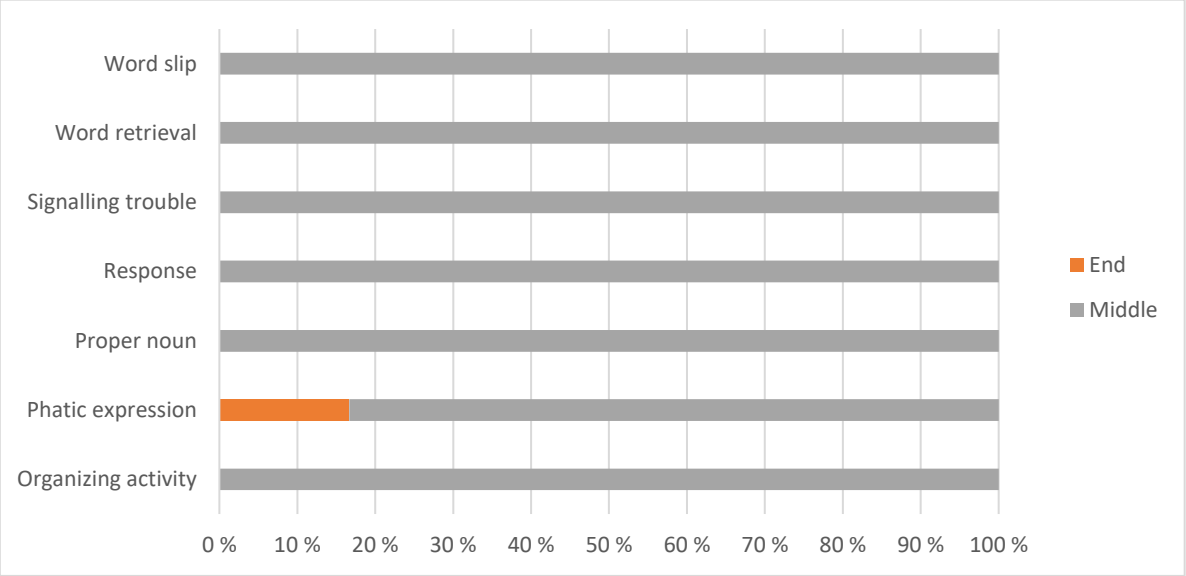

Figure 6. Positions of codeswitches within a task in FUSE.

Interestingly, the FUSE corpus has no instances of codeswitching that occur at the beginning of the task. In fact, all instances occur in the middle of task completion with only one exception of one phatic expression signalling the end of a task. The complete absence of L1 use before or at the beginning of a task also seems to relate to the lack of turn allocation and clarifying instructions in FUSE, which are prominent functions in the HY-Talk corpus and also tend to occur before or at the beginning of a task. Although this is purely conjecture, it is likely that these differences stem from the functions being fulfilled before the recording in FUSE, while in HY-Talk the students made these preparations during the recording. Of course, as discussed in section 4.3, this is a limitation of the data as I am unaware of the circumstances before the recordings.

In the following chapter, I will again go over the results presented in this chapter. The goal is to tie together the results into the very main findings by returning to my initial research questions presented in chapter one and answering them. I will also review my results against the results from previous research on the topic of L1 use in classrooms and in oral proficiency examinations, which I first discussed in chapter 2. By doing this I will find similarities and

differences between my results and those of previous studies and therefore placing my study in the field of studies on L1 use by L2 English students.

5. Discussion

In this chapter, I will review the results of my analysis. I will return to my research questions presented in chapter 1 and answer them according to my findings. Additionally, I will draw comparison between previous research and my results to see how my results compare to previous findings on L1 use in EFL classrooms and oral proficiency examinations.

What are the functions of L1 use in L2 English oral proficiency examinations?

A total of twelve distinct functions for which students used their L1 were found in the two corpora analysed. These functions could more broadly be divided into three main categories: task management, vocabulary, and other. Task management functions work to actively progress the current task in some way. Vocabulary functions were used either for word searches or to signal inbound self-correction of an error. Functions under other are those which did not fit the previous two categories. Two of these categories, response and proper noun, were added for maximum recall's sake and to not needlessly distort the numbers of other functions. Off-task talk, signalling trouble, and phatic expressions seem to serve a more social function rather than an interactional one.

The functions found in the current study closely resemble the results of previous research (Alegría de la Colina & García Mayo, 2009; Storch & Aldosari, 2010; Azkarai & García Mayo, 2015). Previous research also found that the L2 students used L1 in order to manage the task at hand, discuss vocabulary, and for off-task functions. This suggests that the codeswitching students use in oral proficiency examinations closely resembles the codeswitching habits in classrooms. These findings support my initial hypothesis that the codeswitching occurring in L2 English oral proficiency examination situations resembles that of codeswitching occurring in EFL classrooms. In accordance with previous research it was also found that students use L1 in moderation, though this is to be expected in a testing situation. If anything, the variation of functions on my results is somewhat less varied than in previous studies, which is likely due to less varied task types in the corpora. The task modality in both corpora was solely more or less structured oral pair discussions and the students were not asked to produce a e.g. written work. Future research may be interested in the effects of task modality on codeswitching in L2 English testing situations as well as Finnish L2 English classrooms.

The allocation of different functions to broader categories has been done differently in previous studies from my own. Alegría De La Colina and Del Pilar García Mayo (2009) categorised off-task talk as a separate category whereas I have included it under the category of other. I also included the function of alleviating stress by e.g. joking in off-task talk, whereas Alegría De La Colina and Del Pilar García Mayo (2009) categorised it under task management. I believe that the social function of relieving anxiety, while indirectly assisting task management, is not directly related to the task at hand or to the production of the target language. Therefore, I believe this function does not belong under task management. Storch and Aldosari (2010) categorised phatic expressions under task management, whereas I place them under other. I am uncertain why Storch and Aldosari (2010) consider phatic expressions as part of task management functions. As I do not consider them to directly assist task completion, I stand by my categorisation.

The use of the Finnish "eiku" (roughly the Finnish equivalent of the English "I mean") for the function of correction closely resembles the results Nyroos, Sundholm, and Sundqvist (2007) reported for the use of Swedish "eller" (Swedish "or") in English oral proficiency examinations. Both seem to indicate trouble awareness and signal to the co-participant that self-repair is imminent. However, I have doubts about the claims of single particle codeswitches not being accidental. The use of single lexical items I found from my data resembles the automatic code-switching found in Pietikäinen's (2014) study on ELF couples. Pietikäinen (2014) defines automatic code-switching as occurring passively without flagging, hedging, or other explicit markers indicating the switch to be planned or intended (p. 13). According to Pietikäinen (2014), automatic codeswitching is allowed in ELF couple communication because it does not obstruct understanding and is therefore let pass by without much attention (p. 18). This is also the case in both HY-Talk and FUSE, as the students all share a common language, Finnish. Additionally, students have been shown to be aware that they should avoid using their L1 when completing spoken L2 tasks (Storch & Wigglesworth, 2003), which should especially be the case during oral proficiency examinations. I would argue that the use of such single particle codeswitches during oral proficiency examinations is, in fact, accidental. These constructs are so automatic in a student's L1 that they slip by even during production of a foreign language.

Are there any tendencies of the functions and frequencies of L1 use for students with lower grades versus those with higher grades? If so, what kind?

Whether oral proficiency and L1 use have any correlation between them is impossible to determine by these results, but there appear to be some tendencies for different proficiency levels. The results from HY-Talk suggest that, in general, lower proficiency students use L1 more frequently than those of higher oral proficiencies. There also appears to be a tendency among lower proficiency learners to use L1 more frequently for vocabulary functions, while vocabulary functions became rarer as the proficiency grew. Task management functions were more common for higher proficiency learners. This finding on tendencies of L1 functions according to proficiency levels mirrors that of DiCamilla and Antón (2012). The low amount of data from FUSE made the results somewhat indefinite. Even then, the results point to similar conclusion on the functions used by different proficiency levels as the results from HY-Talk; lower proficiency learners used L1 more for vocabulary searches, while task management was used more by students of higher proficiency.

However, it is unclear from these results whether oral proficiency has a direct effect on codeswitching during oral proficiency examinations. As Luoma (2004) points out, oral performance can be affected by many factors such as the task design and the co-participant. This brings to question the extent of the participants' influence on each other's performance. Does co-participant's codeswitching influence the codeswitching of the other student? This subject would warrant further research. Additionally, it is pertinent to keep in mind that the evaluations of the students' proficiency are also subjective and based on their performance on the recorded test alone. One also wonders whether the students' use of L1 during the oral proficiency examination influenced their grades in any way. In Itkonen's (2010) thesis, the evaluators of HY-Talk are reported discussing a student's Finnish hesitation. While it is not made explicit whether this hesitation influenced their grading or not, it seems that the Finnish hesitation was not positively received by the evaluators. From this we can say that codeswitching to a student's L1 during an L2 oral proficiency test, if anything, does not affect the student's grade favourably. However, further research on the possible correlation between codeswitching during oral proficiency examination situations and grading is in order.

Are there differences in L1 use between HY-Talk and FUSE? If so, what kind?

L1 was used quite similarly in both corpora, however HY-Talk featured more varied functions of codeswitching from FUSE. It was also found that in HY-Talk the students used codeswitching more frequently than in FUSE. One possible explanation for this is the difference in size between the two corpora. The FUSE corpus is, as of now, considerably smaller in size compared to the HY-Talk corpus. Therefore, the smaller number of codeswitches is to be expected. Another reason for this difference could have to do with the placement of the codeswitches, as discussed in section 4.5. HY-Talk showed the functions of turn allocation and clarifying instructions most typically appearing at the start of a task, whereas the two functions were completely absent from FUSE. Furthermore, codeswitching in FUSE never occurred at the beginning of a task. This leads me to believe that the students in FUSE came better prepared to the recording while students in HY-Talk had these conversations recorded. Perhaps students recorded for FUSE already had fulfilled the two functions before starting. However, as the circumstances of the recordings are mostly unknown, I cannot say for certain the reasons for this difference.

6. Conclusion

In conclusion, I have established that Finnish EFL students do, in fact, switch to their L1 (in this case Finnish) during L2 English oral proficiency examinations using data from the HY-Talk and FUSE corpora. Additionally, the L1 use during the examinations closely resembles L1 use during L2 English classroom learning tasks as found in previous research. In my data, L1 was used for twelve distinct functions which could in turn be categories under three broader categories: task management, vocabulary, and other off-task functions. The analysis also revealed some tendencies of L1 use for lower and higher proficiency upper secondary school students. Lower proficiency students used codeswitching to L1 for vocabulary functions more than higher proficiency students, while higher proficiency students used L1 more often for task management functions than lower proficiency students. However, due to the small amount of data, further research is in order to make definitive claims of possible correlation between codeswitching and grading. I found that the students recorded so far for the FUSE corpus used codeswitching less than students in the HY-Talk corpus. The variation of different functions for which L1 was used was also smaller in FUSE. These differences are, at least in part, explained by the different sizes of the corpora with the FUSE corpus being considerably smaller than the HY-Talk corpus.

The data for this study is somewhat limited in size, making it difficult to make any indefinite claims about L1 use of Finnish students during L2 English oral proficiency examinations. FUSE especially is, at least for now, quite small in wordcount, which might skew results somewhat. As the FUSE corpus grows in the future, it would be pertinent to conduct further research on the subject. Future research could also concentrate on the possible effect of L1 use on language assessment. The current study unfortunately cannot make claims of the relationship of the two. Even though some tendencies of L1 use could be established for students of different proficiencies, whether codeswitching is an indicator of language proficiency or whether it influences the grading process could not be determined with this dataset.

Previous research has shown students' L1 use to be a valuable pedagogical, cognitive, and social tool for L2 learning. Therefore, omitting L1 completely form the classroom and prohibiting the students from using their L1 is inadvisable. Additionally, outside the classroom students are likely to use English in various ELF communications, in which various

communicative strategies including codeswitching are useful. It is then logical that moderate use of L1 should perhaps not be sanctioned in language testing situations either.

References

Primary sources

The HY-Talk Corpus, English subset. Department of Modern Languages / English and Department of Applied Sciences of Education, University of Helsinki, 2007–2009.

Ehrnrooth, L. (2015). *FUSE – The Finnish Upper Secondary School Corpus of Spoken English*. <https://fusecorpus.eu/about/> [Accessed 22 April 2020]

Secondary sources

Alegría De La Colina, A., & del Pilar García Mayo, M.P. (2009). Oral interaction in task-based EFL learning: The use of the L1 as a cognitive tool. *International Review of Applied Linguistics in Language Learning*, 47, 325–345.

Anthony, L., 2019. AntConc (Version 3.5.8). [Computer Software]. Tokyo, Japan: Waseda University. Available at: <http://www.laurenceanthony.net/software> [Accessed 22 April 2020]

Antón, M., & DiCamilla, F. (1998). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. *Canadian Modern Language Review*, 54, 314–342.

Azkarai, A., & del Pilar García Mayo, M.P. (2015). Task-modality and L1 use in EFL oral interaction. *Language Teaching Research*, 19, 550–571.

Benson, E. J. (2001). The neglected early history of codeswitching research in the United States. *Language and Communication*. 21, 23-36.

Brooks, F.B., & Donato, R. (1994). Vygotskyan approaches to understanding foreign language learner discourse during communicative tasks. *Hispania*, 77, 262–274.

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26, 341–366.

Bullock, B. E. & Toribio, A. J. (2009). *Cambridge handbook of linguistic code-switching*. Cambridge, UK: Cambridge University Press. 1-18.

Dicamilla, F.J., & Antón, M. (2012). Functions of L1 in the collaborative interaction of beginning and advanced second language learners. *International Journal of Applied Linguistics*, 22, 160–188.

Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language Testing*, 26 (3), 423–443.

Ehrnrooth, L. (2017). *Suullisen kielitaidon opetuksen kehittämishanke – VÄISKI.* <https://blogs.helsinki.fi/vaiski-projekti/> [Accessed 11 May 2020]

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge, UK: New York: Cambridge University Press.

Itkonen, T. (2010). *Spoken Language Proficiency Assessment: assessing speaking, or evaluating acting?* (MA Thesis, University of Helsinki) Retrieved from http://urn.fi/URN:NBN:fi-fe201006162058

Lüdi, G. (2003). Code-switching and unbalanced bilingualism. In Dewaele, Housen & Li Wei (Eds.), *Bilingualism: beyond basic principles. Festschrift in honour of Hugo Baetens Beardsmore*. Clevedon: Multilingual Matters. 174– 188.

Luoma, S. (2004). *Assessing speaking*. Cambridge, UK: Cambridge University Press.

Martin-Jones, M. (1995). Code-switching in the classroom: Two decades of research. In L. Milroy & P. Muysken (Eds.) *One Speaker, Two Languages: cross-disciplinary perspectives on code-switching*. Cambridge: Cambridge University Press. 90–112.

Martin-Jones, M. (2000). Bilingual classroom interaction: a review of recent research. *Language Teaching*. 33, 1– 9.

Neokleous, G. (2016). Closing the Gap: Student Attitudes Toward First Language Use in Monolingual EFL Classrooms. *TESOL Journal*. 8, 314–341.

Nyroos, L.; Sundholm, E.; Sundqvist, P. (2007). Code-switched repair initiation: The case of Swedish eller in L2 English test interaction. *Journal of Pragmatics*. 120, 1–16.

Pietikäinen, K. S. (2014). ELF couples and automatic code-switching. *Journal of English as a Lingua Franca*, 3, 1-26.

Rommetveit, R. (1985). Language acquisition as increasing linguistic structuring of experience and symbolic behaviour control. In J. Wetsch (Ed.), *Culture, communication and cognition: Vygotskyan perspective*. Cambridge: Cambridge University Press. 183–204.

Rose, H., & Galloway, N. (2019). *Global Englishes for language teaching*. Cambridge, UK: Cambridge University Press.

Storch, N. & Aldosari, A. (2010), Learners' use of first language (Arabic) in pair work in EFL class. *Language Teaching Research*, 14, 355–375.

Storch, N., & Wigglesworth, G. (2003). Is there a role for the use of the L1 in an L2 setting? *TESOL Quarterly*, 32, 760–770.

Villamil, O.S., & de Guerrero, M.C.M. (1996). Peer revision in the L2 classroom: Social cognitive activities, mediating strategies, and aspects of social behaviour. *Journal of Second Language Writing*, 5, 51–75.