UiT The Arctic University of Norway

Department of Language and Culture

Bilingual vocabulary acquisition

A corpus-based case study

Esther Varhaug

Master's thesis in English Acquisition and Multilingualism

ENG 3991

May 2021

Abstract

A bilingual learning two languages from birth seems to follow the same developmental trajectory as that of a monolingual. However, within vocabulary acquisition persistent and significant differences have been found between bilingual and monolinguals when the languages of bilinguals are compared separately to monolinguals' level. The domain specific nature of a bilingual's vocabularies is an important characteristic in understanding this difference, but also when investigating a bilingual's vocabulary acquisition. The role of context in which a bilingual acquires his or her languages is crucial to the aspect of domain specificity. An important theory here is the Complementarity Principle, emphasizing how different domains in life require different languages and that the vocabularies and proficiencies of a bilingual's languages will develop thereafter. This theory has mostly been investigated in adult bilingualism.

This study examines the English and Norwegian vocabularies of a 2-year-old bilingual girl, Emma. Emma is growing up in northern Norway and has acquired Norwegian and English from birth. The thesis uses corpus data collected between the ages of 2;7-2;11 to examine the expressive lexical characteristics of Emma's vocabularies. Emma is a relatively balanced bilingual, with an asymmetrical code-switching pattern. Her vocabulary levels do not match those of monolingual peers when her languages are compared separately, but when her total and conceptual vocabularies are compared to monolingual vocabulary levels, her results are more comparable. Based on this, Emma's results are discussed in relation to the implications of the complementarity principle and current research on monolingual and bilingual comparisons.

Acknowledgements

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 A corpus study of Norwegian-English bilingual vocabulary

In this thesis longitudinal corpus data is used to explore bilingual first language vocabulary acquisition in light of an influential bilingual language theory, such as the Complementarity Principle (Grosjean, 1997). Transcriptions of recordings from a child who is acquiring Norwegian and English from birth is analyzed to investigate expressive language characteristics and their relation to the theoretical implications and research on vocabulary acquisition in this area. The corpus data is of a simultaneous bilingual 2-year-old, Emma, who acquires English (as her heritage language) and Norwegian (as her societal majority language) from birth (Bentzen 2000). Using current research in the area, I examine different characteristics of Emma's expressive language skills to investigate different aspects of her vocabulary and her use of and separation of Norwegian and English. This is then compared to current research results in the area and discussed to which degree implications of the influential complementarity principle can be met in preschool children's vocabulary acquisition. In the area of vocabulary acquisition persistent monolingual and bilingual differences have been found, leading to discussions on which factors are influential in bilingual vocabulary acquisition and how best to measure the distributed characteristics of the bilingual vocabulary. There are several factors influencing vocabulary acquisition. Language characteristics such as translation equivalents and code-switching are also explored concerning their relationship to vocabulary. Exposure patterns and language dominance are also aspects that will be explored. Vocabulary measures such as total vocabulary and conceptual vocabulary will be counted, and the results will be compared to those of monolingual peers acquiring English and Norwegian. This is done in order to explore quantitative similarities and differences based on different measures of the bilingual vocabulary in accordance with previous research within this field.

1.2 Thesis structure

In this thesis an overview of previous research in bilingual first language vocabulary acquisition is presented in Chapter 2. Here some notably early studies are presented concerning bilingual and monolingual comparisons before the different concepts, issues and research evidence are reviewed. This will be followed by an introduction of the complementarity principle before similarities and differences in Norwegian and English language are addressed. Chapter 3

presents the background, research questions and methodology of this study. Methods for data analyses and a discussion of some of the potential methodological issues are provided. In chapter 4 results from the data analyses are presented according to the research questions. These results will then be further discussed in relation to research within the area and the complementarity principle in chapter 5. Chapter 6 concludes this thesis, briefly summarizing the findings and placing them within the research area of bilingual first language vocabulary acquisition.

2 Child bilingualism and early bilingual vocabulary acquisition

2.1 Bilingual first language acquisition: A brief overview

Research on bilingualism has demonstrated that when and how you acquire a language, or moreover, two languages, can strongly influence the development and fluency of that language or languages. Because of this research often divides between simultaneous and sequential bilingualism to more clearly separate and narrow down differing backgrounds and conditions (Chondrogianni, 2018, De Houwer, 2009). Simultaneous bilingualism is usually defined as bilingualism where the two languages are introduced early in life. Here research usually draws the line around the age of 3 – 4, as this is an important milestone in the language acquisition where most of the structures like nouns and verbs in the first language is relatively set, along with some prosodic features of that language (Chondrogianni, 2018). After the age of 3 – 4, researchers usually talk of sequential bilingualism or child language 2 learning (CL2) (Chondrogianni, 2018). Another definition used is bilingual first language acquisition (hereafter BFLA), which is a term used to describe backgrounds where two languages are introduced from birth (De Houwer, 2009). The key criterion behind this definition is that both languages are present from birth, so that order of first appearance and timing of first appearance of either language is not an issue of possible influential factors. In this thesis the BFLA definition will be used.

Research in this area within the last decades has also demonstrated how simultaneous acquisitions of two languages from birth, like BFLA, in several aspects resemble that of monolingual first language acquisition. Here it has been demonstrated that early child bilinguals have the same fundamental milestones in language development across the domains of

language as monolinguals. Oller et al (1997) demonstrated how bilinguals before the age of 1 engaged in canonical babbling at the similar ages of onset as the monolinguals. Later bilinguals start demonstrating word comprehension and word production within the same timeframe as monolingual children (De Houwer, 2009). Other research demonstrated that bilinguals also acquired grammatical features of their languages within the same timeframe as monolinguals (Hoff et al, 2012). BFLA then has been demonstrated to represent an instance of first language acquisition, as" …the development of each of the bilingual's languages proceeds in the same way and leads to the same kind of grammatical competence as in the respective monolingual children." (Meisel, 2006, p.95). At the same time, within vocabulary or lexical development significant and persistent differences between monolingual and bilingual children have been found (Unsworth, 2013; Bornstein, Putnik & De Houwer, 2006).

2.2 Bilingual vocabulary acquisition

Vocabulary acquisition is the process of mapping words to concepts. Words are perceived through phonological representations and a crucial part of vocabulary acquisition is to map these representations on to object representations, also called concepts. A lexical network is thus built through the developmental process of connecting these mappings with semantic meaning. Thus, more than one mapping appears to happen and the process of acquiring a vocabulary is a multifaceted one, developing over time. Research on vocabulary acquisition in the preschool years has been an active area in BFLA in recent years building on an extensive body of research on monolingual first language acquisition (hereafter MFLA) (e.g., Bialystok, Luk, Peets, & Yang, 2010; David & Wei, 2008; Pearson, Fernandez & Oller, 1993; Saffran, Werker & Werner, 2006). Early research on BFLA often focused on the comparison of vocabulary in monolingual and bilingual acquisition.

One frequent finding when investigating and comparing early bilingual and monolingual vocabulary levels, was how bilingual children consistently scored lower on both receptive and expressive vocabulary scores (Unsworth, 2013). In their study Hoff et al (2012) found that when comparing the Spanish-English bilingual children between the age 1;10 to 2;6 with English speaking monolingual peers it was demonstrated how monolingual children's vocabularies were significantly larger and demonstrated greater vocabulary gains over time. This was when comparing the languages of the bilinguals separately. In another study, English-

Spanish bilingual children in the ages 2- 3 were also demonstrated to score lower than monolinguals on expressive vocabulary skills when they were tested in English language skills to compare predictability of parent and teacher reports on other vocabulary measurements (Vagh, Pan & Mancilla-Martinez, 2009). The consistent quantitative difference in vocabulary levels between bilinguals and monolinguals led to the assumption that there might be an academic disadvantage for bilingual children, as vocabulary is a predictor of academic achievement (Bialystok et al, 2010) Furthermore, it made it difficult to be able investigate potential language disabilities in bilingual children. However, when vocabularies were investigated and compared it was based on monolingual norms. Receptive and expressive standardized test, built on the monolingual norms, were used as measures. These measures would investigate the vocabulary of only one of the bilingual's languages, and in addition, compare it to that of a monolingual. This led to the question of to what degree a bilingual's vocabulary levels can be expected to reach that of a monolingual, as a bilingual's language learning is spread across two languages.

An influential study regarding this issue was the study of Bialystok et al (2010). Their study was a large-scale study where they investigated receptive vocabularies in 1.738 children between the ages 3 – 10. All the children spoke English at school, in kindergarten or other situations, whilst speaking a non-English language at home. When testing the children, they used the standardized receptive vocabulary test Peabody Picture Vocabulary test and found that that the monolinguals outperformed the bilinguals at every age point they were tested. The authors point out that this was not unexpected as a bilingual divide their language learning across their two languages. What they additionally explored was the contextual distributional differences between vocabularies of monolinguals and bilinguals. Here they discovered that when they compared vocabularies belonging to the school context, the vocabulary levels between monolinguals and bilingual were comparable. Where the difference between the two groups was, was in the home context, which in the bilinguals' case, was not surprising as all the bilingual children in the study spoke a different, non-English, language at home.

Bialystok et al. (2010) conclude that "the smaller vocabulary for bilingual children in each language is not an overall disadvantage but rather an empirical description that needs to be taken into account in research designs, especially in tasks that involve verbal ability or lexical processing. Moreover, the vocabulary deficit for home words in English in the bilingual children is almost certainly filled by knowledge of those words in the non-English language,

making it likely that the total vocabulary for bilingual children is in fact greater than that of monolinguals" (Bialystok et al, 2010, p. 530).

These findings emphasized the distributed characteristics of bilingual vocabularies as something that should be considered when investigating bilingual children's language levels. As the authors point out, the words that they did not know in English, they probably knew in their home language. Because they use those words at home, with their family, and do not need them in their everyday school life, their vocabularies are, to a large extent, a result of use and need of that langue in particular contexts. Thus, the study highlighted a crucial aspect of not only bilingual vocabulary acquisition, but also bilingual language development in general. Bilingual children's vocabularies are distributed, i.e., they are specific to the context in which they are needed and used. Thus, it has been proposed that domain-specificity is the characteristic feature of vocabulary in child bilinguals. In the following section, I present the domain-specificity proposal as well as the complementarity principle.

2.3 Domain specificity and the complementarity principle

An influential theory on the domain-specificity of bilingual languages is the complementarity principle (hereafter CP) (Grosjean, 1997). The CP as principle emphasizes how "Bilinguals usually acquire and use their languages for different purposes, in different domains of life, with different people. Different aspects of life require different languages." (Grosjean, 1997, p.165). This theory has several important implications for how one regards bilingual language acquisition in general, as it heavily emphasizes the role of the context in which the languages are learned. Depending on the social context, different languages would be needed. Through growing up in multilingual settings, different areas of life, i.e., domains, such as family vs school/work, could call for the use of different languages. Thus, a bilingual's social contexts and interlocutors are crucial influencers of exposure, language use patterns and consequently language fluency. One language might be used in fewer domains and with fewer people, whilst the other is used more often. As a result, imbalances between the languages can develop, depending on the number of different domains in which the bilinguals can hear and use their languages. Additionally, the theory also has implications for the distribution of the vocabularies. Not only will differing numbers of domains influence amount of exposure for each language, differing number of domains in which the languages are used will also influence

which words are required for different domains. This would lead to different contents of the vocabularies. Another important implication of the CP is how the relationship between the languages becomes dynamic. Language fluency is something that can change over time, across domains and interlocutors depending on use and exposure. In this way the CP emphasizes the role of the context and domains in everyday life of a bilingual as a way to understand the relationship of fluency and dominance between majority and heritage language.

To further explain the dynamic relationship between the languages, not only over time, but also across situations, the CP emphasizes the presence of a language mode continuum. There are different levels of a bilingual's language modes, where a mode is the level of activation of the bilingual's languages and language processing mechanisms. According to the CP the activation happens along a continuum, where bilinguals can, depending on situations and interlocutors, find themselves along different points. In the one end of the continuum there is a monolingual language mode – a context and/or conversation where only one language is appropriate – and thus where one or the other language is activated whilst the other is deactivated. Speakers can also be at the other end of the continuum, where they are in a bilingual mode. This happens when they are in a context and/or conversation where they can use both languages. Here both languages are activated, however usually one of the languages is the main communicative language and thus also more active than the other language. Bilinguals can also find themselves at different points along this continuum. This has implications not only for the language used, but also for level of activation of both languages and therein also activation of crosslinguistic influences. Being in a bilingual mode will open for more use of code-switching. This phenomenon is discussed in detail in section 2.5.4.

A study by Chiaro (2009) demonstrated implications of the CP. Here 39 female and 20 male participants were interviewed on their linguistic habits. All participants were bilingual and had been in a bilingual relationship for an average of 10 years. The participants were couples from Europe, the US and Canada, involving 24 different language pairs. It was investigated which language participants used in a variety of social domains, in addition to which language they used as a couple. Across various domains the languages used would differ, where they would predominantly use their partner's mother tongue with the partner's relatives. With their own relatives, they would predominantly use their own mother tongue. Additionally, they would predominantly use their own mother tongue in domains such as prayer and worship, whereas situations and domains that were shared between partners showed several different language-pattern uses. Here language use showed great variation between which language was

used, but also a higher amount of both languages used (for full overview see Chiaro, 2009). This variation in language use across domains and interlocutors demonstrate the main argument of the CP; bilingual "…use their languages for different purposes, in different domains of life, with different people. Different aspects of life require different languages." (Grosjean, 1997, p.165). (For more research around the CP see Grosjean, 2015). However, research into the implications of the CP has focused mainly on adult bilinguals or school-aged children (Montanari et al, 2018). The extent to which the implications of the CP can be related to early child bilingual vocabulary acquisition still needs to be explored.

In sum, the CP thus emphasizes the importance of the contextual background regarding the distributional characteristics of a bilingual's vocabulary that are relevant to my investigation of early child vocabulary acquisition. Vocabularies for the different languages can vary according to need and use of each of the languages. The two vocabularies are not two completely separate entities, rather there are crosslinguistic influences. This again has further implications in relation to what needs to be considered when investigating bilingual vocabulary acquisition:

(1) Depending on the characteristics of the domains in which bilinguals find themselves, their vocabularies can be overlapping or separated. In domains where both languages can be used or activities are the same/similar, they may know vocabulary for that specific domain in both languages; what is referred to a shared domain. Lexical overlap is defined in terms of translation equivalents (TEs), which I explore in section 2.3.2.

(2) Different language modes will result in different levels of activation of the languages. This will influence levels of code-switching, something that will also be addressed in a section 2.3.3.

(3) Differing numbers of domains in which a bilingual is exposed to the languages and in which the languages are used will influence level of fluency in each language. This leads to the possibility of one language being more dominant than the other. This will be explored in the following section.

2.3.1 Vocabulary, language dominance and the role of input

Because a bilingual might be exposed to the two different languages in different amounts, an imbalance between the vocabularies might occur. In instances where one language is needed across more domains and between more interlocutors than the other, the CP predicts an imbalance, leading to one language becoming dominant. Input and dominance have been

investigated as influential factors in early bilingual vocabulary acquisition along with multiple intertwined factors, including the role of the language(s) in society, amount and quality of the input as well as input patterns from primary caregivers (Eilers, Pearson, & Cobo-Lewis, 2006; Byers-Heinlein, 2013; Bail, Morini & Newman, 2015).

The amount of input has long been demonstrated to be an influential factor in vocabulary acquisition in general (Brandeker & Thordardottir, 2015). In their study Brandeker & Thordardottir (2015) investigated language exposure in relation to vocabulary and performance on nonword repetition tasks in 60 children in the age range of 2;5 to 3;6. Here children had differing levels of exposure to the languages in question; English and French. They found that amount of exposure correlated significantly with both receptive and expressive vocabulary. However, this correlation was stronger for expressive vocabulary. In addition, they found that for bilingual children who had spent 40% or more of their time in an English language environment scored as a group within normal range of vocabulary and grammar levels to that of monolinguals.

It is the input from primary caregivers that will encompass the initial building block of the bilingual's language acquisition, as it is from this input they start their building of phonetic inventories that they later build vocabularies from. What has been an aspect of some focus in this area is the input model selected by parents in a bilingual setting, something that will also influence amount of exposure in each language the bilingual child will face initially (De Houwer, 2007).

An influential study looking into these factors was performed by De Houwer (2007). Here 1899 families were investigated for the relationship between minority/majority languages spoken by the parents at home and the children's minority language use. It was demonstrated that the pattern of language(s) spoken by the parents at home correlated with language use of the children. A frequently used model is the one parent- one language approach. This approach, however, did not always achieve a desired result of a bilingual child's language abilities (De Houwer, 2007). According to De Houwer (2007), in families where this strategy was employed at least one of the children would end up speaking the majority language only in 12 % of the cases. In other cases, other parental input strategies have been employed, where both parents speak their minority or heritage language at home, a language that is not used by society. Even if one or both parents know that language as well. With this approach, DeHouwer found that only in 3 % of the cases would one of the children end up speaking majority language only.

Where the minority language was spoken by both parents at home, or where one of the parents used minority and the other used minority and majority language, the children had higher rates of minority language use. Whereas the one-parent-one language strategy in several cases did not provide sufficient input for the child to use the minority language. This was also found in a study by Yamamoto (2001). They studied language use in families of one native Japanese-speaking parent and one native English-speaking parent. Out of the 209 families that participated, 61 used the strategy of native English speaker speaking only English and the other parent speaking both. 45 of the children from this approach spoke English to their native English-speaking parent. Out of the 46 families that employed the one parent-one language approach, 25 of the children used English with their native English-speaking parent. This also demonstrates a smaller percentage of children using their minority language through the one parent – one language approach then when both parents use the minority language. Where the author remark: "In other words, the more that both parents use the minority language and the less that the minority language parent uses the mainstream language in speaking to the child, the more likely that the child will use the minority language to the parent who is a native speaker of it." (Yamamoto, 2001, p. 102). Here the authors also question to what extent the expectation of language use from the parents also influences child minority language use. The strict adherence to one or the other languages might sub-textually communicate expectations to the child about which language it is appropriate to use. This is something that will be further discussed in the section on code-switching. Parental language use pattern will also be addressed in the investigation of Emma's expressive language skill, as her parents employ the minority language only in the home.

The role of the heritage language in society also plays an influential role. It can play an influential role regarding attitudes towards the minority language. Attitudes towards a language may influence willingness to use the language. This will, however, often be more an existent factor for older children (Eilers, Pearson & Cobo-Lewis, 2006, Yamamoto, 2001) and will therefore not be more elaborated here. But it can also play an influential role regarding how many domains where the minority language is used that is accessible to the child. In societies where a heritage language is rare, it naturally involves fewer domains in which the child will encounter the heritage language, and as mentioned amount of exposure is an important aspect of vocabulary acquisition (Brandeker & Thordardottir 2015, Hoff et al, 2012). It has been demonstrated how even balanced bilinguals change that their language use patterns when

entering school and gain more exposure to majority language. As a result, the majority language develops and the heritage language stagnates (Montanari et al, 2018).

Amount of exposure does not only influence whether the bilingual child will use the minority language or not, it will also influence the balance between the languages (Yip & Matthews, 2006). Even if a bilingual child will use the minority language, it does not necessarily follow that they will speak this as well as the majority language. Asymmetry in language proficiency is not unusual when a child has received more exposure to one of the languages (Paradis & Nicoladis, 2007). This asymmetry is often referred to as language dominance. Language dominance entails that the bilingual is more proficient in one of the languages or displays a more advanced development in one of the languages (Paradis & Nicoladis, 2007). Dominance has had demonstrated effects on vocabulary acquisition. In their study on 1;6 and 2;5-year-olds, Hoff et al (2012) demonstrated the expected gap in vocabulary levels when comparing the languages of bilinguals separately to those of monolinguals. They further demonstrated how the bilingual children that were English dominant had vocabulary scores that were more comparable to monolingual levels, than were score from Spanish dominant bilinguals. In addition, dominance also exert an influence on lexical activation, where words are recognized faster in the dominant language (DeAnda, 2016).

In addition, the role of dominance is dynamic, as it potentially changes over time as amount of exposure varies. This has been demonstrated by Montanari et al (2018), where 126 Russian-German bilingual children in the age span 6 – 10 were investigated for changes in exposure affected vocabulary. The children were beginning school and thus were increasing their exposure to the majority language, German. They demonstrated a stagnation in the development of the vocabulary in their heritage language, whereas their vocabulary in the majority language increased (Montanari et al, 2018). In addition, the overlap between the languages increased. As well as the children demonstrating an increase of words that were exclusive to the majority language, as well as a decrease of words that were exclusive to the heritage language. These children were older than the age span investigated in this thesis. However, the start of school represents a transition into a domain with intensive exposure to the majority language. This thesis addresses the potential similarity in change of domains as a potential influence on Emma's vocabulary overlap in English and Norwegian.

Thus, language input patterns, amount of exposure and dominance are important factors in early bilingual vocabulary acquisition and contribute to the distributed characteristics of bilingual vocabularies that will be explored in the present study.

2.3.2 Translation equivalents and their influence on early vocabulary acquisition

Translation equivalents (TEs) can be defined as word pairs from the different languages, with formally distinct word forms, that can be considered to have an equivalent meaning in a particular context (De Houwer, 2009). An example would be the Norwegian word *bil* and the English word *car*.

There have been several different studies on TEs in bilinguals, where bilingual children as young as 8 months display presence of TEs. The overview of these studies and their results are summarized in in Table 1 (Bosch & Ramon-Casa, 2014). TEs are acquired before children reach the 50-word milestone (Bosch & Ramon-Casa, 2014). The amount of TEs reported across different studies varies, but it is clear from Table 1 that TEs represent between 20% to 40% of the expressive vocabulary of bilinguals before the age of 3.

Table 1 Overview of translation equivalents (TEs) in the early lexicons of bilinguals acquiring different pairs of languages (Bosch & Ramon-Casas, 2014).

Authors	Languages	N	Method	Ages	% Translation Equivalents (TEs)
Pearson, Fernandez, & Oller (1995)	Spanish-English	2 7	MacArthur CDI	0;8–2;6	31% (90 words)
Deuchar & Quay (2001)	Spanish-English	1	Diary and recordings	0;10–2;3	33% (27% in 50 words)
Nicoladis & Secco (2000)	Portuguese-English	1	Audiovisual recordings	1;0–1;6	20%–25% (18 months)
Holowka, Brosseau-Lapre´, & Petitto (2002)	French-English	3	Audiovisual recordings	0;7–2;2	27% (50 words)
	French-LSQ	3			29 %
Junker & Stockman (2002)	German-English	1 0	Language Development Survey (LDS)	2;0–2;3	43.7% (range 10.7–84.2)
Schelleter (2002)	German-English	1	Audiovisual recordings	1;11–2;8	> 30%

| David & Wei (2008) | French-English | 1 3 | MacArthur CDI and recordings | 1;0–3;0 | Up to 40% linked to language dominance |

Note. LSQ: Langue des Signes Que´becoise; CDI: MacArthur-Bates Communicative Development Inventory.

TEs have been demonstrated to have a facilitative effect on word learning, where knowledge of one word in one language helps acquisition of that word in the other language (Bosch & Ramon-Casas, 2014, Schelletter, 2002). Furthermore, Byers-Heinlein & Werker (2009, 2013) demonstrated that bilinguals who knew more translation equivalents used disambiguation as a strategy much less than did bilinguals who knew fewer translation equivalents in word learning tasks. Disambiguation refers to a word learning strategy often found in monolingual vocabulary acquisition, where children will most often assume that a novel word naturally belongs to a novel object. Because of their early encounter with TEs bilingual children are more likely to accept that an object can be referred to with different names and are therefore less likely to use disambiguation as a word learning strategy (Byers-Heinlein & Werker, 2013). This emphasizes TEs a facilitator of vocabulary acquisition. Not only do TEs demonstrate a facilitatory effect on acquisition, but they also demonstrate a facilitative effect on word recognition and lexical access, where words with a known TE are faster accessed and retrieved than words without a known TE (Poulin-Dubois et al, 2013, Poulin-Dubois et al, 2018). This, as I will discuss in the results section on code-switching, can have implications on code-switching utterances.

Differing amounts of overlap have also been connected to specific properties of the languages (Bosch & Ramon-Casas, 2014). How close or distant phonological forms between words of the specific languages in question are, has been demonstrated to influence acquisition of TEs (Bosch & Ramon-Casa, 2014). This was indicated in a single case study by Schelletter (2002) where a 2-year-old German – English child demonstrated a more balanced use for form-similar TE nouns across languages, than for TE nouns that were dissimilar. By form-similarity it is here referred to word pairs that were similar in meaning and phonological form. In addition, for form-similar nouns, the gap between when they were used was much smaller, indicating that they were more easily acquired. This was also supported by Bosch & Ramon-Casas (2014) finding that phonological form proximity between words across bilinguals' two languages facilitates early lexical acquisition. The more similar the languages are, the more they will facilitate acquisition of words that are similar across the two languages. This means that for

languages such as for example Spanish-Catalan, Norwegian – Danish, TEs can represent an early acquired and relatively large proportion of child's vocabularies. As mentioned earlier, this might influence vocabulary measures such as total conceptual vocabulary, as different language pairs might then influence amount of overlap. Children with similar language pairs might demonstrate a larger overlap and therefore a smaller conceptual vocabulary score than children whose languages are very different from each other. In this case, it might underestimate the sum of the child with a similar language pair's vocabulary knowledge.

Thus, TEs are an important feature of young bilinguals' vocabulary as this is present in their vocabularies from an early stage in acquisition and seem to have a facilitative effect on word acquisition.

2.3.3 Code-switching

Code-switching is a common occurrence in bilingual communication and entails that words or grammatical properties are transferred from one of the languages over to the other, resulting in mixed utterances (Ribot & Hoff, 2014). Here it is necessary to highlight that there are different ways to look at and define code-switching. Where some research subsumes all types of mixed utterances into the definition of code-switching (Yow, Tan & Flynn, 2018), there is also research where a distinction between code-switching and code mixing is made. Code-mixing is sometimes used for the mixing of two languages within the same sentence, either at a lexical level or a syntactic morphological level and is discussed to be an unconscious process. Code-switching is also used for the mixing of the two languages but may occur within as well as across sentences and discussed to be a conscious process where the mixing supports communication (Yow, Tan & Flynn, 2018). In this thesis code-switching will be regarded as all instances of mixed utterances, as the focus will be on the amount of and pragmatic use of code-switching.

There are differing results in investigations on code-switching in relation to vocabulary acquisition, where some studies have demonstrated it have a negative effect on vocabulary acquisition, whereas others have found no such effects on vocabulary acquisition (Byers-Heinlein, 2013, Bail, Morini & Newman, 2015, Kaushanskaya & Crespo, 2019, Byers-Heinlein, Morin-Lessard, & Lew-Williams, 2017). Though there are differing results as to the effects of code-switching on vocabulary acquisition, another aspect of studies on code-

switching has often been on the occurrence of and pragmatic use of code-switching ad what it communicates about the bilingual's linguistic competence (Yow, Tan & Flynn, 2018).

An overview on code-switching studies by Nicoladis & Genesee (1997) found that bilingual children as young as 1,5 years old and up to the age of 6 uses code-switching seemingly as a communicative strategy in as much as they use it to fill lexical gaps, for instance when they do not have the appropriate translation equivalent. This has been a widely accepted explanation and been demonstrated in some studies (for overview see Nicoladis & Genesee, 1997) and would also concur with the implications of the CP (Grosjean, 2015). They additionally found how amount of intra-utterance code-switching used ranged from 0% up to 45%, with there being no relationship between amount of code-switching and age of the children. This was demonstrated through the age at which the lowest amount code-switching occurred was 2;2 years and the age at which the highest amount occurred at age 2;1. Also, in the case of inter-utterance, a similar pattern was found, where amount of mixing ranged from 0% up to 70%. However, the authors point out that variation could be due to differing definitions of code-switching.

As code-switching is demonstrated to be so variable across participants and ages, the question then becomes how code-switching relates to vocabulary acquisition. The implication of the CP is that code-switching is a strategy to aid communication when vocabularies are distributed across domains and words are "missing" in a specific situation because words a certain topic or activity is acquired in the other language. It acts as a communicative support, especially in the case of one of the languages being dominant. Having a more developed vocabulary in one language should predict support from the stronger, dominant, language into the weaker language. Additionally, the language mode demanded by the situation should also influence amount of code-switching that occurs. In a bilingual mode the CP emphasizes how both languages are more activated, which makes words from both languages more readily available. This nonselective activation of both languages has also been demonstrated in research concerning lexical activation (Von Holzen & Mani, 2012). As such the discourse strategy chosen by the parents could be predicted to influence language mode for the children. In the paragraphs below two case studies will be explored to see how two different factors, language dominance and parental discourse strategies, influence code-switching behavior in young bilinguals.

Language dominance is, as mentioned earlier, in a strong relationship with vocabulary acquisition, where English-dominant bilingual speakers of English and Spanish demonstrated comparable expressive vocabulary levels to that of monolinguals (Hoff et al, 2012). Asymmetry in levels of proficiency has frequently been found to correlate with asymmetry in occurrence of code-switching between the languages (Paradis & Nicoladis, 2007). Here code-switching of the dominant language into the less proficient language occurs more than from the weaker language into the dominant language. The question that has further become important regarding investigations on code-switching, is what role the code-switching plays. In a study by Nicoladis and Secco (2000) the role of code-switching as a communicative strategy for missing words was investigated. They investigated the code-switching in a Portuguese-English bilingual child from the age 1;0 to 1;6. It was found that that the child code-switched more when speaking in his weaker language. In addition, he code-switched to fill lexical gaps in the less proficient language in up to 90% percent of the code-switching instances. Here the authors emphasize how the lack of appropriate translation equivalents, cross-language synonyms, to be the main drive behind respective lexical gaps. The results from this case study seem to emphasize that a larger vocabulary in one language influences what type of code-switching behavior is elicited in the child: a lexical gap-filling for words not known on the weaker language. An important question regarding this study is what happens if the languages are balanced? An implication here would be that there would be far less code-switching, as the more balanced the languages are, the more TEs are likely to be present according to David & Wei (2008). This issue is of special interest for the present study, as I will show in the results section bilingual Emma seems to be rather balanced in her two languages, Norwegian and English. Nicoladis and Secco (2000) also discussed how code-switching behavior by the parents could potentially influence code-switching in the child. The parents used code-switching as a communicative strategy to meet the child's limited linguistic resources by either code-switching with the translation equivalents the child lacks or by switching to use the words that the child already knows in the other language.

The discourse strategies used by the parents as an influence on code-switching behavior in the child have also been investigated. Lanza (1992) studies a Norwegian-English bilingual child in the age range of 2;0 to 2;7 focusing on the child's use of code-switching and how the context, in terms of the parents' strategy when code-switching occurred, potentially influenced the code-switching. The child demonstrated a predominance of function words in the code-switching, something the author highlighted to be instances of language mixing, as grammatical

elements are not usually mixed as a communicative strategy, nor is a 2-year-old aware of grammatical categories. The directionality of the child's transfer, using Norwegian grammatical morphemes in both English and Norwegian code-switching utterances but no English grammatical morphemes into Norwegian, was interpreted to demonstrate language dominance. In this respect investigating whether Emma's code-switching consists of function words as an instance of language mixing and not instances of lexical gap-filling, could inform on the distribution of her vocabularies. In addition, it was investigated the sensitivity to interlocutor in terms of how much code-switching occurred with each parent. The parents used a one parent-one language approach. It was demonstrated that the child had no problem separating the languages when communicating with the different interlocutors. What was additionally found was that the parents employed different strategies, that also elicited different code-switching patterns from the child. Whereas the mother employed a minimalist grasp strategy, a strategy negotiating a monolingual language context, the father employed a move on strategy, negotiating a bilingual context. The author emphasized the different strategies to elicit differing degrees of code-switching utterances (Lanza, 1992).

Code-switching according to the CP is a communication strategy connected to the domain-specific vocabulary distribution that result from amount of exposure, language dominance and parental discourse strategy, among other factors. Language dominance and parental discourse strategy are influential factors regarding code-switching as they potentially influence amount and directionality of code-switching. As such these studies highlight different aspects of code-switching relevant to Emma's language background, who has a minority monolingual language context at home, but speaks the majority language, Norwegian, outside of home and also during the recordings with her investigator. However, as we will see Emma is a relatively balanced bilingual. Will there be any instances of code-switching in her utterances, and if so, what role do they play? Are the lexical items supporting communications where a translation equivalent is lacking, or are they function word morphemes indicating a mixing of the dominant language into the weaker language like in the Lanza study? Are there any instances of parental discourse strategy that might influence her code-switching behavior? These and other questions will be addressed.

2.4 Measures of early bilingual vocabulary acquisition

Recent research on child bilingualism, including vocabulary acquisition, emphasizes the problem with single-language assessment of bilingual children (e.g., Core et al.2013). Single-language assessment refers to testing one language of a bilingual child and comparing the results with those of monolinguals for that language. According to Core at al. (2013) and many others, this practice is problematic, since bilingual children may have lower vocabulary scores than their monolingual peers when they are tested in only one of the languages, because a bilingual child's vocabulary knowledge is distributed across two languages, and a single language assessment captures only part of what a bilingual child knows (Hoff et al., 2012; Junker & Stockman, 2002; Pearson, Fernández, & Oller, 1993; Thordardottir et al., 2006; Vagh, Pan, & Mancilla-Martínez, 2009). Therefore, two measures of bilingual children's combined vocabulary have been proposed: conceptual vocabulary and total vocabulary. According to Core et al. 2013, total vocabulary is the sum of the words a child knows across two languages and conceptual vocabulary gives the child credit for knowing concepts rather than words, and concepts that are represented in both languages are counted only once. For example, if a child says *bil* in Norwegian and *dog* in English, she gets credit for one concept, even though she is able to produce two different word forms.

2.4.1 Total vocabulary

When both languages for bilinguals are investigated and accounted of, i.e., total vocabulary knowledge, bilinguals do not seem to possess smaller vocabularies than their monolingual counterparts (Hoff et al, 2012). In their study, Hoff et al (2014) investigated expressive vocabulary level trajectories from 22 months to 48 months in 31 monolingual English-speaking children and 26 Spanish-English bilingual children. They also found that when languages were compared separately, the bilinguals scored lower than the monolinguals. However, when total vocabulary score of the bilinguals was compared to that of monolinguals there was no significant differences between them. This was also demonstrated in Core et al (2013), where 47 Spanish-English children from 22 to 30 months were tested for mean vocabulary scores and mean growth in their total vocabulary. It was demonstrated how the total vocabulary scores of the bilingual children were not significantly different from the single language scores from the monolinguals. This demonstrates that when total vocabulary is measured, bilingual children might demonstrate similar vocabulary knowledge as

monolinguals. The authors concluded that total vocabulary could be used as a measure of early vocabulary development.

Total vocabulary is a widely used measure of bilinguals' vocabulary knowledge and will be used here to investigate Emma's vocabulary levels compared to those of monolinguals in the same age range. These results demonstrate the importance of measures such as total vocabulary as it will include all the words a bilingual knows and sidestep issues regarding domain specificity possibly limiting vocabulary knowledge for bilinguals. What has been suggested to be a possible limitation to this method, however, is the fact that total vocabulary scores can overestimate bilinguals' vocabulary knowledge. More specifically, getting a score for each word, even if they both represent the same concept, will overextend their vocabulary knowledge (Core et al, 2013). Other researchers have pointed out that bilinguals should get a score for each word. They emphasize that there are phonological learning processes behind all words. Therefore, all words should also be accounted for (Core et al, 2013).

2.4.2 Conceptual vocabulary

As total vocabulary has been questioned to overestimate the vocabulary knowledge of bilinguals, several studies additionally employ a measure of total conceptual vocabulary. Total conceptual vocabulary involves counting the number of concepts the child knows, regardless of which language it is expressed in and giving only one score/number for that concept. This would evade the problem of overestimating the vocabulary knowledge of bilinguals. However, here research has demonstrated differing results when conceptual levels are compared with monolingual levels. In their study Core et al. (2013) found that the conceptual vocabularies of the 47 Spanish-English participants were not only significantly smaller than the total vocabularies, but they were also growing at a slower rate than the total vocabularies. This was also found when the conceptual vocabularies were compared with monolingual vocabulary levels at age 30 months. Interestingly, this was not the case for the ages 2;2 and 2;5, where conceptual vocabularies did not differ significantly from the monolingual vocabulary levels.

In a study by Junker & Stockman (2002) the conceptual vocabularies of 10 English-German speaking bilinguals from 2;4 to 2;7 years old were not significantly smaller than those of monolingual peers. However, in a study by Thordardottir (2006) 8 French-English bilingual in the age range 2;6 and 3;0 demonstrated significantly smaller conceptual vocabularies when

compared with monolingual vocabulary levels. Differing results could seem to indicate that as a measure conceptual vocabulary is not without limitations.

One limitation, as argued by Quick et al (2018) and Core et al (2013), is how the use of conceptual vocabularies can be limiting the bilingual language knowledge in as much as it relies to a large extent on the finding of translation equivalents. Translation equivalents (TEs) can be defined as word pairs from the different languages, with formally distinct word forms, that can be considered to have an equivalent meaning in a particular context (De Houwer, 2009). An example would be the Norwegian word *bil* and the English word *car*.

Both Core et al (2013) and Quick et al (2018) emphasize that it is a limitation to rely on the count of TEs in conceptual vocabulary. When conceptual vocabularies are scored, a concept that is represented in both languages only gets counted as one concept known. This is problematic they emphasize, because far from all translation equivalents are isomorphic, i.e., has a direct translation into another language. For instance, the Norwegian word "kylling" (chicken) can refer to both the edible noun "Hühnchen" in German or the noun for the small yellow baby chicks "Küken". Furthermore, the amount of overlap in a bilingual vocabulary may vary according to language pairs learnt (Smolak et al, 2020, Bosch & Ramon-Casas, 2014). This will be further elaborated on in the methodological issues section.

2.5 Previous research on Norwegian-English bilingual children

Several of the before mentioned studies are quantitative investigations. For my study corpus data is used. In this section a short introduction on similarities and differences between Norwegian and English will be presented before an overview of studies on Norwegian-English bilingual preschoolers that also used corpus data will be provided.

As English and Norwegian are both Germanic languages, there is substantial superficial structural and lexical overlap with respect to word order and vocabularies. Both English and Norwegian have a basic SVO word order (Anderssen, M. & Bentzen, K., 2018). Even longer sentences between English and Norwegian show remarkable similarities. There are other grammatical similarities, too. Like the possessive adding of an 's at the end of the words. There are of course differences as well, like the adding of an apostrophe in English. Or how cases of subject-verb agreement in English are somewhat more complicated than that of Norwegian. In

terms of words and vocabularies, there are also several similarities in the form of cognates. Cognates can be defined as words where the form and meaning of words from two different languages considerably overlap. This overlap can be due to similarities between languages or the result of a borrowing process (Otwinowska, A. & Szewczyk, J.M., 2019). Examples of cognates between Norwegian and English respectively are words such as *baby-baby, mann-man, te – tea*. However, with many cognates, there are also many false friends, or false cognates. Here the words are formally similar to words in the other langauge, the meanings, however, differ between the languages (Otwinowska, A. & Szewczyk, J.M., 2019). There are also a lot of these between Norwegian and English respectively, like the words *bare-only, bra-good, hell-good luck, men-but.*

Studies on Norwegian-English BFLA are few, and some of them are syntactically focused, not necessarily lexical. But some have found results regarding phonological inventory development in congruence with earlier mentioned research on bilingual perceptional development. In a case study by Johnson, C.E. & Lancaster, P. (1998), a Norwegian-English bilingual boy demonstrated a separation of phonological inventories in Norwegian and English at the age 1;9. At the age of 1;9 his vocabulary consisted of 126 Norwegian words and 125 English words, 35 of which were identified as translation equivalents, where an additional 19 words were indeterminate due to difficulty of understanding his pronunciation. In addition, he quickly started to separate language use by interlocutor. In the Lanza (1992) study the 2;0-year-old Norwegian-English bilingual girl distinguished between the languages according to context and interlocutor, and as such showed ability to code-switch. Interestingly, her investigation of the instances of code-switching demonstrated differences between grammatical and lexical mixing between the languages. Whereas her grammatical mixing was attributed to language dominance, her lexical mixing was of a more complicated nature, as her mixing pattern did not corroborate with her dominance pattern, and her use of lexical items could not be attributed to lack of translation equivalents. Leading the author to emphasize that there are other factors involved in lexical code-switching than language dominance and lexical gaps. She here emphasizes what the interaction between interlocutors allows for, where by one parent a monolingual context was negotiated through indications that mixed responses were inappropriate responses, whereas the other parent negotiated a bilingual context by moving the conversation forward ignoring instances of code-switching and thus "allow" for code-switching. This indicates that how code-switching is met by interlocutors might influence amount of its occurrence. Thus, interlocutor sensitivity is present in even young bilingual

toddlers. Lanza's study thus demonstrates that lexical code-switching need not necessarily demonstrate a lack of translation equivalents, rather it could also be conditional to discourse strategy by parents. This will have implications for the role of the CP, where code-switching is emphasized to be used as communicative support. Even though Norwegian and English are lexically similar and could influence overlap, so could parental-discourse strategy.

Another relevant study was conducted by Walla (2017), who used corpus data to investigate code-switching in a Norwegian-English bilingual girl between the ages 2;3 to 3;3. Using recordings that were transcribed in the CHAT format, she analyzed conversations between the mother and the child to find code-switching patterns and what this could say about code-switching correlates such as language dominance, language modes and discourse strategy. Through analyses of her MLUs and conversing habits she finds how code-switching happens predominantly in the girl's heritage language, English, while in her majority language Norwegian, little code-switching occurs. In addition, English code-switching was predominantly lexical words, whereas Norwegian code-switching was predominantly function words. Which is attributed to her language dominance for her majority language, Norwegian. This language dominance is attributed to the girl's language context, where she is predominantly met by the majority language. Her parents follow the one parent-one language approach; however, they do speak Norwegian with each other. This entails that she is predominantly exposed to her majority language and, as the author highlights, she has rare occasions to find herself in a monolingual English language mode. Additionally, when she is speaking her minority language, the move-on strategy is often employed, where Norwegian responses are allowed, leaving her in a bilingual language mode. As we will see Emma has a different language background and relationship between the languages, which is set in relation to her different code-switching pattern and discussed regarding other research results in this area. These types of analyses demonstrate the importance of being able to use corpus data.

Regarding the Lanza and Walla study, what lexical items are prevalent in a vocabulary is also relevant. It has been shown that bilingual children's vocabularies have a larger number of nouns than verbs (Golinkoff & Hirsch-Pasek, 2008, Gentner, 1982, Childers & Tomasello, 2006, Bornstein et al, 2004). Here research has demonstrated that linguistic features of the target language (Waxman et al, 2013, Qiu & Winsler, 2017), as well as influence of mothers' language use and the social context are important contextual aspects of word learning (Harris, 2004). This prevalence has often been attributed to concepts of objects being "…perceptually and

conceptually more stable, and therefore more readily acquired, than concepts of actions or events, which involve relations among objects." (Waxman *et al*, 2013, p. 156). This, however, need not be the case, as other research has demonstrated an effect of language properties and input affecting this relationship between object words and action words, where a larger amount of action words in input will result in a larger portion of the vocabulary to consist of action words (Gopnik & Choi, 1995, Tardif, 1996). Where Asian languages have more focus on actions in their language and where verbs are used more in their daily language exert an influence on vocabulary composition so that action words make out a much larger portion of the vocabulary than action words (Qiu & Winsler, 2017). As such context again make for a crucial aspect of vocabulary acquisition.

The corpus data used in this study is from a small Norwegian-English BFLA child named Emma. Earlier studies have also investigated Emma's languages. In her thesis, Bentzen (2000) investigated verb placement and definite DPs as evidence of systemic grammatical mixing in Emma's language use patterns. Through analysis of Emma's verb placement and definite DPs she found that Emma demonstrated crosslinguistic interdependency, transfer, in her syntax and thus that Emma's languages did not develop autonomously. In addition, she also found this transfer to occur in both languages, a bidirectional patten of code-switching, where a higher rate of switching happened in her English. This was found to be indicative of other driving forces behind Emma's code-switching than language dominance alone and it was suggested that Norwegian might be Emma's strong language. In a later corpus-based study, Emma and two other Norwegian-English bilingual children' languages were used to investigate acquisition of residual verb second and to which extent crosslinguistic influences affected these structures (Anderssen & Bentzen, 2018). The age range between the three children were 1;6 – 3;9. Here the three different the children exhibited different patterns in the use of the constructions. The authors suggested that different usage of the construction was not necessarily due to language dominance, but a result of possible interpretations of an ambiguous English language system. Especially when this ambiguous language system met with a consistent system, such as the Norwegian V2.

The studies mentioned above are all instances of Norwegian-English bilingual language studies, focusing on different aspects of preschool bilingual language acquisition. Instances of code-switching patterns in relation to language background and syntactic development and transfer. However, none of them mainly investigate vocabulary acquisition. In this study the

corpus data from Emma will be used as an example of young, preschool BFLA to investigate vocabulary acquisition.

3 Present study

The starting point of the thesis is the Complementarity Principle emphasizing the role of the context in language acquisition and use, where language skill depends on the need of that language in different domains and with different people. The implications of the CP are that level of fluency, hereby meant how proficient a speaker of a given language is, and language dominance depend on the need of that language and that it will be domain specific. Domain specificity in this context entails that different areas, or domains, of life require different languages, vocabularies etc. Furthermore, the communicative competence of a bilingual, different in nature than that of monolinguals, "…makes use of one language, of the other, or of the two together (in the form of mixed speech) depending on the situation, the topic, the interlocutor, etc." (Grosjean, 2015, p. 68). The need of that language in different situations and domains will then exert an effect, or influence, on vocabulary development as little need of a language will entail little exposure and/or use, which further will influence development of that language. Furthermore, vocabulary characteristics, such as code-switching and translation equivalents, which are normal characteristics found in bilinguals' languages, are also influenced by CP implications. The role of code-switching is to support communication in domains or topics where one of the languages has developed a smaller role and thus needs the help from the dominant language in that domain or topic to communicate effectively and what the context allows for of language mixing (Grosjean, 2015). Here constraints on code-switching would also take form through where along the language mode continuum the bilinguals find themselves in, where again context is crucial. The implications for translation equivalents would here entail that the role of the domain would influence amount of overlap between vocabularies.

Research looking into the CP has generally used data from adult bilinguals or school-aged children. The degree to which an adult bilingual's lexicon(s) and that of developing toddlers are similar is being researched. This project attempts to explore the CP and its implications in relation vocabulary acquisition in young bilingual children. It will be explored through the expressive lexical knowledge in a young simultaneous bilingual first language acquisition

setting, something that has not yet been investigated thoroughly. Here Corpus data from an English-Norwegian bilingual 2-year-old girl, Emma, with a simultaneous bilingual language background, is investigated to look into aspects of expressive language skills. These aspects comprise dominance, lexical item quantity and distribution, code-switching and the role of translation equivalents in relation to the development and use of a bilingual's vocabulary. In addition, this is also compared to the vocabulary of a Norwegian monolingual and an English monolingual in order to investigate the distributional characteristics of bilinguals' vocabularies similarities and differences looked into by previous research. The corpora of the bilingual and monolingual participants are presented in more detail in section 3.2.

3.1 Research questions

Based on the project aims and the corpus data available, the following research questions will be explored in this thesis:

3.1.1 RQ I: What are the quantitative characteristics of Emma's expressive language skills in the heritage language, English, and the majority language, Norwegian?

The characteristics I will be looking into are whether she uses the same quantity of lexical items in both languages. By lexical items I have here focused on nouns and verbs. I will also explore the quantitative distribution of the nouns and verbs. Other characteristics that will be investigated are her total and conceptual vocabularies. Code-switching and translation equivalents are also characteristics but will be addressed as separate research questions.

As for her total and conceptual vocabulary, recent research debate which of the two measures best represent the bilingual's total language knowledge as they explore different aspects of the lexicon's composition (Core et al, 2013). Total vocabulary is usually larger than conceptual vocabulary, as cross-language synonyms only account for one score. Therefore, it is predicted that her total vocabulary score is larger than her conceptual score, as it is predicted there be some amount of overlap due to the similarities of her languages. This because TEs, cross-language synonyms, are demonstrated to have a facilitatory effect in vocabulary acquisition, especially for similar languages (Schelletter, 2002, Bosch & Ramon-Casas, 2014) The CP would also predict a larger total vocabulary, as it emphasizes distributed

vocabularies across domains. As for the conceptual vocabulary, however, the predictions here depend on whether the domains are classified as separate or shared. If the domains are separate, a small overlap is predicted, as the vocabularies will be for two separate domains. This again will predict a larger conceptual vocabulary. However, due to the playful nature of kindergarten attendance, playing and eating, a relatively large amount of overlap could be expected as these are activities she also practices at home. This could influence the outcome of her conceptual vocabulary. If her activities are similar to those of home, a large amount of overlap can be expected, which would entail a perhaps smaller conceptual vocabulary, as only one word per concept is counted. This would also be predicted by the CP, where shared domains usually result in larger overlap of vocabularies and thus indicating that her conceptual vocabulary would be influenced by this. Methodological issues regarding the count of overlap between nouns versus verbs are discussed.

Research on lexical items have found differing results, where the majority emphasize a noun prevalence in young bilingual's lexical composition, like that of monolinguals (Correia & Flores, 2017). As both English and Norwegian are languages whose structural features are related to a relative salient and frequent use of nouns and verbs in the input that in turn influences degree of the noun advantage (Gentner, 2006), Emma's languages are expected to follow the noun prevalence in her lexical composition. Here CP predictions are difficult, as this is a relatively unexplored area for the principle. On the one hand, her lexical composition is predicted to result from her exposure to the two languages in the different domains, which seems balanced, where, however, one might expect a slight advantage to her Norwegian, due to her kindergarten attendance. This because she spends most her days here, with several different people, which the CP would then emphasize to be a "stronger" domain. On the other hand, she has spent at least her first year of life with her home, and thus English, would be the "strong" domain.

3.1.2 <u>RQ II</u>: Are there instances of translation equivalents in Emma's vocabularies, and if so, to what extent?

Research looking into translation equivalents, also called doublets, has found them to be a prevalent and important aspect of bilingual language acquisition (De Houwer, 2009, Legacy et al, 2015). Not only have they been attributed to demonstrate how the bilingual child can separate two linguistic systems (Nicoladis, 1998), but they have also been demonstrated to

form an early part of vocabulary acquisition and to grow in accordance with the expanding vocabulary (Bosch & Ramon-Casas, 2014, Legacy et al, 2015). In addition, they also make out a relatively large proportion of the bilingual vocabulary, where research has found that TE's can, on average, make out up to 30 % of a bilingual's vocabulary (Legacy et al, 2015), though this varies individually. In can further be influenced by language pairs, where the similarity between languages can facilitate acquisition of TEs (Bosch & Ramon-Casas, 2014). Norwegian and English as similar languages could influence amount of TEs. Thus, it is expected that there will be an amount of overlap between Emma's languages, not only in object words, nouns, but also in action words, verbs. Due to individual variations in language acquisition, how many TE's one can expect is difficult. However, the CP will here originally predict that as her heritage language is spoken at home, where she will then be in monolingual mode, and her majority language is spoken in Kindergarten, also in a monolingual language mode, there might be a relatively small overlap between words. On the other hand, due to the playful nature of kindergarten attendance, playing and eating, a relatively large amount of overlap could be expected as these are activities she also practices at home. What could be questioned, however, is the degree to which specific toys overlap compared to actions in playing. This could influence number of nouns versus number of verbs that would be overlapping. Here, again, if her home domain and kindergarten domain are instances of such shared domains, her overlap is expected to be relatively large.

3.1.3 RQ III: Are there instances of code-switching in Emma's utterances and if so, what role do the code-switching instances play?

An active area in research is when and why children code-switch, with multiple possible explanations (Ribot & Hoff, 2014). Two large areas of research have been to look into asymmetries between the languages, i.e., dominance, and context. Contextually research has found correlations to exposure and discourse strategies (Lanza, 1992). Here the role of the languages in different social contexts also influences language choice. Children are sensitive to the situation, to which degree the situation can be seen as bilingual or monolingual, as well as the "expectations" of the interlocutors (Lanza, 1992, Yow, Tan & Flynn, 2018). Such pragmatic differentiation has been demonstrated in bilingual children as young as 1-year old (Nicoladis, 1998). This has also been investigated in relation to language dominance, where research has demonstrated more use of code – switching when expressive skills in one of the

languages is weaker than the other (Ribot & Hoff, 2014). This, then, suggest code-switching can also have a function of filling lexical gaps. However, research has also found that children code-switch even if they know the words in the other language (Quick et al, 2018). Thus, both context and language dominance are influential factors when it comes to vocabulary acquisition and to code-switching.

Based on previous research the question then becomes to what extent instances of code-switching from Emma seem to be due to filling of lexical gaps. Are instances of code-switching utterances predominantly functor words indicating a mixing of languages or content words indicating lexical gap filling. As her social context is mainly monolingual, this should also influence her code-switching amount according to the CP. She talks English with her parents, and Norwegian with the investigator. This would also be predicted according to the CP, where instances of code-switching and borrowings should be to a small degree, as her languages are used in separate domains and her language mode, according to the bilingual continuum mode, is monolingual in both domains. If there are instances informing of discourse strategies from her parents, this could also influence where Emma would find herself along the language mode continuum. Thus, unless discourse strategies by her interlocutors call for a bilingual mode, any instance of code-switching should be at a minimum and be of some communicative supportive intent and purpose.

3.1.4 <u>RQ IV</u>: Do the characteristics reflect the differences in Emma's language exposure and language use patterns? Are her vocabularies distributed (home vs. daycare)?

Exposure, both parental and societal, has been found to be very influential factors in vocabulary acquisition (Hoff et al., 2012). Here a study by Thordardottir (2019) found how amount of exposure seem to be a stronger influence than timing, age of exposure, in language performance in school-aged bilinguals when comparing simultaneous and sequential bilingual children. However, both quantity and quality of input influences rates of development (Hoff & Core, 2015). With Emma being exposed to her heritage language at home and her majority language in her kindergarten, her languages could be expected to be somewhat balanced as these are two domains that make up a large proportion of her daily life. As for language use, there are several factors that can influence language use as well, as mentioned in previous

research questions. Norwegian is the majority language, to which she is exposed in multifaceted settings, kindergarten, social interactions outside her home, media. Socio-politically English does not have an official role in the social context, outside her home, except for music, television, and other social media. This makes her exposure to English reduced in number of domains. However, as the parents have adopted a minority language use at home, she has been predominantly exposed to English in her first year of life. Concerning exposure and language use patterns, this could influence Emma's vocabulary development and size and language use patterns in the longer run. English's role in society is not expected to play an influence on her language pattern yet, as her awareness of this relationship at the age of two is not expected. What here most likely is an influence on her language pattern use is her parents' distinction between interlocutors and situations, something children are sensitive to (Lanza, 1992, Yow, Tan & Flynn, 2018), so her language use patterns are expected to be along the monolingual language continuum according to the CP. The vocabulary distribution should, according to the CP's language continuum modes, be expected to be separate, however, due to the playful nature of kindergarten, and the possibility of these domains being shared, her vocabularies might not be separated as such.

3.1.5 <u>RQ V</u>: How do Emma's expressive lexical skills compare to that of monolingual peers in each language?

It is well documented that, when languages are measured separately, bilingual vocabularies are often smaller than those of their monolingual peers (Unsworth, 2013, Thordardottir, 2011, Smithson, Paradis & Nicoladis, 2014). Measures used sometimes yield different results, with total vocabulary often being equal to that of monolinguals, whereas total conceptual vocabulary have yielded equal or smaller vocabularies than that of the monolingual peers (Hoff et al, 2014, Core et al, 2013). This would lead to the expectation of Emma's number of nouns and verbs, in conceptual count, to be smaller than those of her monolingual pers, whereas her total vocabulary is predicted to match those of the monolinguals'. Developmentally research has found that bilinguals develop at the same milestones and develop the same categories at the same age as monolinguals (De Houwer, 2009), so Emma's characteristics, presence, and distribution, of nouns and verbs are expected to match those of the monolinguals as well as her MLU's are expected to be on the same level as those of the monolinguals. Here the CP will predict her separate vocabularies to be smaller to that of her monolingual peers, as her

vocabularies are spread/separate to different domains, whereas her total vocabulary should be more comparable.

3.2 Method

3.2.1 Corpus data analyses

To investigate the research questions this thesis uses corpus data. Corpus data is generally defined as analyses of datasets of machine-readable text that has been transcribed with indexes and tags so that a reading program, as a tool, such as CLAN, can allow for a large-scale reading, searching, and manipulating of linguistic information (McEnery & Hardie, 2012). Within linguistics, though it might be useful in other research areas as well, using corpus data is incredibly useful within a range of different areas, including, but not limited to areas such as language learning, sociolinguistics, discourse analysis, theoretical linguistics (McEnery & Hardie, 2012). Regarding child language acquisition, this is a field with an advantage when it comes to the use of corpus data, mainly due to CHILDES. CHILDES is a databank of collected transcriptions from different studies involving child-speech that is open and free to use for different studies. Here transcriptions are made in a CHAT format, which allows the tool program CLAN to run frequencies in a very user-friendly way. The use of corpus data allows for both qualitative and quantitative investigations, especially for data amounts that are difficult and/or time consuming to perform by hand. Corpus data and tools such as CLAN thus allow for reliable information on frequency lists, listing words that appear and number of occurrences. For instance, instead of having to manually count number of verbs and nouns occurring in transcriptions from recorded sessions, the right transcription of such recorded data will allow for a tool to extract the necessary information needed and list the different words and the frequency with which they occur. This allows for investigation of vocabulary, for instance, both quantitative and qualitative aspects thereof. It can, through differing angles of frequency, give insight into patterns of words and phrases, which a lot of different theories within the field of psycholinguistics build on (McEnery & Hardie, 2012). These investigations do not only investigate the children's own language, but also parental input, which is an important factor in language learning. In addition to a high reliable method of counting frequencies, there is also the high replicability of results procured from such an investigation method. It is thus a useful tool for language observation that can, together with other data collection methods, help investigate questions that are not easily measured or countable in other respects.

As any other research method, however, it is not without its limitations. Being a large-scale qualitative and quantitative tool for looking into frequency information, it naturally limits what research questions can be answered. To a certain extent there are also limitations when it comes to transcriptions. Though luckily in child language linguistics the CHILDES databank and transcription method of transcribing into CHAT format is relatively set and universal, there is, however, the process of transcribing child speech, which can be a laborious task in addition to being difficult, due to pronunciation differences and other modes of speaking. There is also the matter of finding suitable transcriptions to run analyses on, for comparisons between children the background information should hopefully match for age, socioeconomic status, length of recording, MLU's to have as similar research context as possible. What is also an aspect of analyzing the data is how, in the instance of English for example, the CLAN tool does not necessarily separate nouns and verbs that share a form, for instance *fall* or *dance,* is it the verbal *to fall, to dance* or the nouns *a fall, a dance.* hence the investigator must go through instances of such words to check the use of the words in the sentences.

3.2.2 Bilingual and monolingual corpora in this study

This investigation uses corpus data collected by Bentzen (2000) from Emma, a two-year old English-Norwegian bilingual girl. Emma was audio recorded for a duration of about an hour per session over the age span from 2;7.10 months till 2;10.9 months. Recordings in English were made during playtime and other family situations, such as mealtimes, with family members. Her recordings in Norwegian were made during playtime with the investigator (Bentzen, 2000). Emma spoke Norwegian with her investigator and to begin with she did not know that the investigator could speak English. Regarding language background English is the spoken language, by both her parents, at home in Norway. Norwegian is the language spoken at her kindergarten, which she started attending around age 1, and by her Norwegian grandparents. Here the mother is American and the father Norwegian, however, they all speak English in the home. The parents follow a strict adherence to the minority language use only and that they do not code-switch (Bentzen, 2000). This is supported through examples in the dialogue demanding monolingual language modes. There are a few words they use resolutely in Norwegian, such as *matpakke* (lunch), *barnehage* (kindergarten) and the word *sånn* (there/like this).

This means that Emma has not been equally exposed to English and Norwegian from birth. Until she has started kindergarten, she has only been exposed to Norwegian through grandparents and other situations, where her parents have spoken Norwegian with friends (Bentzen, 2000). This means that Emma is aware that her parents know both languages. However, a refusal to use Norwegian if her parents have tried to address her in Norwegian further demonstrates that they seem adhere to the monolingual minority language pattern (Bentzen, 2000). Regarding vocabulary acquisition, this background is interesting as it allows for investigation into amount of exposure. She had a predominantly English exposure up until the age of one. At this point she started attending kindergarten and has received a large amount of exposure to Norwegian. How this language background manifests in her language levels through MLUs and vocabulary levels between the two languages would give a strong indicator of whether amount trumps timing, as has been demonstrated in school-aged bilinguals (Thordardottir, 2019).

For the English monolingual comparison, data from a North American English monolingual child was sampled from the CHILDES databank. The recordings were from the Brown corpus, of a boy in the age range from 2;2.16 till 3;2.2. Here subsamples that best matched the bilingual's age at recordings and MLU's were selected for comparison. The background of the English monolingual child was also from a middle- to upper-middle-class professional family, where recordings were made on a weekly basis and had a duration of about an hour. As for the Norwegian monolingual child, corpus data was provided for, where samples that best matched the bilingual's age and MLU's were used.

3.2.3 Data analyses

The files from the data collection were all in CHAT format and the data was analyzed through CLAN. Here data on Mean Length Utterance (MLU), as a measure of overall language development and language dominance, and Type/Token Ratio (TTR), as a measure of lexical diversity in their utterances, were collected in addition to calculations of total and conceptual vocabularies. Total number of words used, and total number of different words used were also collected from the CLAN analyses for each recording. Nouns and verbs are words that mainly make up the vocabulary and communicative basis for very young children as adults would categorize them (De Houwer, 2009). As such, nouns and verbs were counted and used as an indication of vocabulary for this thesis. Total vocabulary and conceptual vocabulary was

counted from the nouns and verbs collected from the corpus data. Total vocabulary was counted by adding total number of different nouns and total number of different verbs used for each language and adding these two numbers together. Conceptual vocabulary was counted by first finding the overlap between the languages and counting these as instances of one concept before adding concepts occurring in only one of the languages. Here there were some methodological issues that will be addressed under the methodological issues section. An additional analysis was undertaken to account for frequency of nouns and verbs. As the Emma CHAT files had been used in an investigation on syntactic features, nouns and verbs were not marked in transcription. Because of this nouns and verbs were sorted and counted manually. This was done in an EXCEL spreadsheet, where other words than verbs and nouns were eliminated from the list, except for in code-switching instances. Two different lists, one for nouns and one for verbs, were created for each recording Here proper nouns were included as well, as some of the proper nouns were names referring to playful objects, such as Winnie the Pooh etc. Furthermore, when counting through totals of different nouns and verbs, plural forms of nouns and verb conjugations were excluded as they are morphosyntactic aspects of the language, rather than lexical semantic aspects. In cases when it was not clear whether it was a noun or a verb, the dialogue was checked for intended meaning. The CLAN analyses would list how many times a word was used, these numbers were used to find total number of verbs and nouns used. An additional manual count was performed to find total number of different verbs used and total number of different nouns used. After going through all the recordings for all three children this way, a count of total number of different words was calculated.

In addition, degree of overlap in terms of translation equivalents was investigated. Here when speaking of overlap between words, it is referred to number of translation equivalents. A manual count and list for nouns and verbs that were overlapping between Emma's English and Norwegian was made for each recording. The spreadsheets were used to compare nouns and verbs occurring and find the overlap. These were summarized in total number of different verbs overlap and total number of different nouns overlap. One last expressive lexical characteristic that was analyzed was amount, use and directionality of code-switching, as this is a usual feature in bilingual communication. One that, however, can shed light on both language use pattern and relationship between the languages. Code-switched words were taken from the lists and put together in a separate Excel spreadsheet, here all lexical instances of code-switched words were included.

3.2.4 Potential issues with methodology

As with most methodologies there are limitations to procedures. Mean Length of Utterance is generally used as predictor of children's expressive language and is often also used as a measurement to explore differences in linguistic productivity, which can also assess dominance in bilinguals' language settings (Quick et al, 2018). Here an MLU was calculated for both her languages, as the Corpus data was collected with the languages separated for the different recordings. However, mentionable is that utterances with code-mixing were not excluded from the MLU calculations, which could potentially influence results of her MLU (Quick et al, 2018). Particularly in her English, where there is a substantial amount of Norwegian code-switching. Eliminating these utterances from the MLU analysis could potentially yield a different picture of her expressive language skills, as has been found for other studies where they have separated code-switching utterances from monolingual utterances when analyzing MLUs (Quick et al, 2018). Here it was found that the highest MLUs were for code-switching utterances. If her high amount of Norwegian code-switching in her English were subtracted, this could potentially give different picture of the relationship between her languages, as it could indicate a different dominance picture than what has been found in the preset study. This could also have relatively influential consequences for the remaining analyses where her language balance is related to other language characteristics.

A further issue was encountered when counting overlap. As discussed, a problem when finding translation equivalents and using them in the count for conceptual vocabulary is to which degree the TEs are isomorphic (Core et al, 2013). This problem was encountered here as well, especially with regards to overlap between action words. The English verb *to get,* for example, can refer to having something given to you, but it can also mean that you will go and find something that is currently not present. Which Norwegian translation equivalents it will count as, as it can be translated as *å få* or *å hente*, respectively, will depend on the situation. Here only one of the translation equivalents was counted. This does mean, however, that there might be language knowledge that will not be accounted for.

Recording situations is a valuable method of collecting data. However, the degree to which a complete language profile can be drawn from one-hour recordings is also questionable, as language use may simply be subject to how talkative the child was that day. Though here multiple recordings over time should increase the likelihood of encompassing several aspects of her language knowledge. Also, this is a case study. Case studies are invaluable in the

possibilities of in-depth analyses and studying. However, as there is large individual variability, conclusions are drawn very tentatively and without any assumption of generalizability to other bilingual children. Hopefully, the results and conclusions mentioned can possibly highlight aspects to be explored further or that can back up on already established findings.

4 Results

4.1 RQ I: What are the quantitative characteristics of Emma's expressive language skills in the heritage language, English, and the majority language, Norwegian?

4.1.1 Emma's MLUs and TTRs

Emma's MLU of the different languages, as seen in table 2 and 3, demonstrates her MLU's to be relatively similar for both Norwegian and English, with slightly smaller MLU's in some of her Norwegian recordings.

Table 2 MLU in Emma's English

Rec. no.	Rec. 2	Rec. 4	Rec. 6	Rec. 8	Rec. 10	Rec. 12
Age	2;7.14	2;8.5	2;8.17	2;9.2	2;9.23	2;10.8
MLU	**4,06**	**4,53**	**4,15**	**4,31**	**4,65**	**4,37**

Table 3 MLU in Emma's Norwegian

Rec. no.	Rec. 3	Rec. 5	Rec. 7	Rec. 9	Rec. 11	Rec. 13
Age	2;7.21	2;8.7	2;8.20	2;9.11	2;9.25	2;10.9
MLU	**3,72**	**3,94**	**4,94**	**3,92**	**4,13**	**4,39**

Here we can see the lowest MLU for her in English to be 4,06 and her highest to be 4,65, whereas her lowest MLU in Norwegian is 3,72. Interestingly, Emma's highest MLU in Norwegian is 4,94, indicating that her expressive language skill in Norwegian can be very high. Here again conclusions are drawn tentatively as speech activity can vary between recordings for other reasons than language skill. With an average MLU of 4,17 in Norwegian across recordings and an average of 4,35 in English across recordings, the difference of her MLUs is not big, a slight dominance in English over Norwegian, and she can therefore be regarded as a relatively balanced bilingual.

Comparing her Type/Token Ratios (TTR) in Table 3, indicate that her English is slightly more lexically diverse. Here her lowest TTR is 0.156 and her highest 0.371. According to these numbers, the amount of different words she uses varies greatly too. In her Norwegian her lowest is 0.133 and 0.232 is the highest TTR, indicating smaller lexical diversity in her Norwegian language use pattern than her English. The total TTR across all recordings also demonstrates a greater lexical diversity in her English, with an average of 0.244 TTR in English, whereas the average TTR in Norwegian is 0.179. This could perhaps demonstrate an aspect of her vocabulary, that she possibly has a smaller vocabulary in Norwegian. Although her MLU for Norwegian indicate that she knows how to make as complex sentences as in English, it is possible that she might not have as many words as in English, as the TTR demonstrate the extent to which the same words are being used over and over or whether words used are varied.

Table 4 *Type/Token ratios (TTRs) in Emma's corpus*

Ages (English & Norwegian)	English	Norwegian
2;7.14 & 2;7.21	0,282 (186/660)	0,194 (169/869)
2;8.5 & 2;8.7	0,156 (257/1647)	0,232 (149/642)
2;8.17 & 2;8.20	0,205 (280/1364)	0,164 (219/1332)
2;9.2 & 2;9.11	0,192 (245/1275)	0,184 (208/1132)
2;9.23 & 2;9.25	0,259 (207/798)	0,17 (226/1331)
2;10.9 & 2;10.8	0,371 (99/267)	0,133 (279/2099)

The age is for both recordings, where the digit after main age corresponds to English and Norwegian age at recordings, respectively.

4.1.2 Quantity and distribution of lexical items, i.e., nouns and verbs

4.1.2.1 Quantity of nouns and verbs

If we look at Emma's languages separately, see figure 1, we can see a clear prevalence of nouns over verbs. In English, Emma has 204 nouns in her total count of different nouns across all recordings, whereas her count of different verbs shows 114 (for a complete data overview see appendix 1A and 1B). Her Norwegian has a similar pattern, with 183 different nouns across recordings and 96 verbs. Within both languages then there is a clear majority for nouns over verbs. With a difference between English and Norwegian nouns on 21 and 18 on verbs, there is a small imbalance in quantitative distribution of these lexical items.

Figure 1. *Total number of different nouns and verbs in English and Norwegian*

	English	Norwegian
■ TDN	204	183
■ TDV	114	96
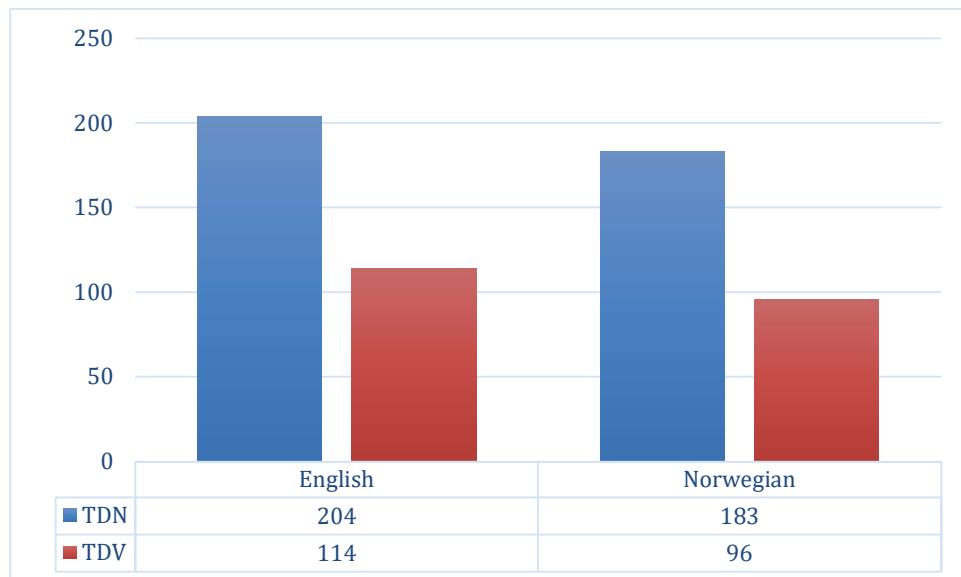
TDN= total number of different nouns, TDV= total number of different verbs

4.1.2.2 Distribution

When looking into her distribution of lexical items first her total number of nouns used (TNN) and her total number of verbs (TNV) used were compared with total number of words (TNW) used in recordings. Here it is evident from table 4 how verbs are used more often in her expressive language than nouns.

Table 5. *Total number and total number of different nouns and verbs in Emma's corpus.*

	Table 4 Total number and total different number of nouns and verbs used							
	English				Norwegian			
Ages	TNN	TDN	TNV	TDV	TNN	TDN	TNV	TDV
2;7.14-21	86	43	178	34	111	38	223	37
2;8.5-7	208	65	360	45	67	31	165	30
2;8.17-20	292	78	288	51	113	43	287	38
2;9.2-11	144	53	292	47	79	48	249	46
2;9.23-25	64	51	186	41	99	52	296	46
2;10.9-8	34	14	50	16	156	72	472	48

TNN= total number of nouns,TDN= total number of different nouns, TNV=total number of verbs, TDV= tital number of different verbs.

This goes for both English and Norwegian. This could possibly be explained through how action words are explanatory in nature and highly relevant as to what is going on around and/or with an object in use. In addition, multiple different action words can describe a situation around the same object. Such as the example in (1), where there are five different action words around the same object: train:

(1)   I got a long train.   2;7.14

        (N: Jeg har et langt tog)

        I got a new train.

        (N: Jeg har et nytt tog)

        now is mine turn.

        (N: nå er det min tur)

        now I'm gonna +...

        (N: nå skal jeg…)

        like to have mine train fixed.

        (N: vil ha toget mitt fikset)

        I need to fix mine train.

        (N: Jeg må fikse toget mitt)

It should also be questioned, however, the extent to which these are the same action words over and over. In example (1) there were different action words used. In example (2) we can see how the same action word is used repeatedly.

(2)   sånn og så **tar** vi den der inni og **tar** lokket på.   2;9.11

        (E: There, and then we take that one in and put the lid on)

        og så **tar** vi den here der inne.

        (E: and then we take this one in here)

        tar vi den på.

        (E: we put it on)

og (.) **tar** vi der inni her og **tar** den der på sånn.

(E: and we take it in here and put that one on like this)

Therefore, an additional comparison was made, where use of number of different nouns (NDN) and verbs (NDV) was compared, see figure 2.

Figure 2. Distribution of total number of different nouns and verbs across recordings.
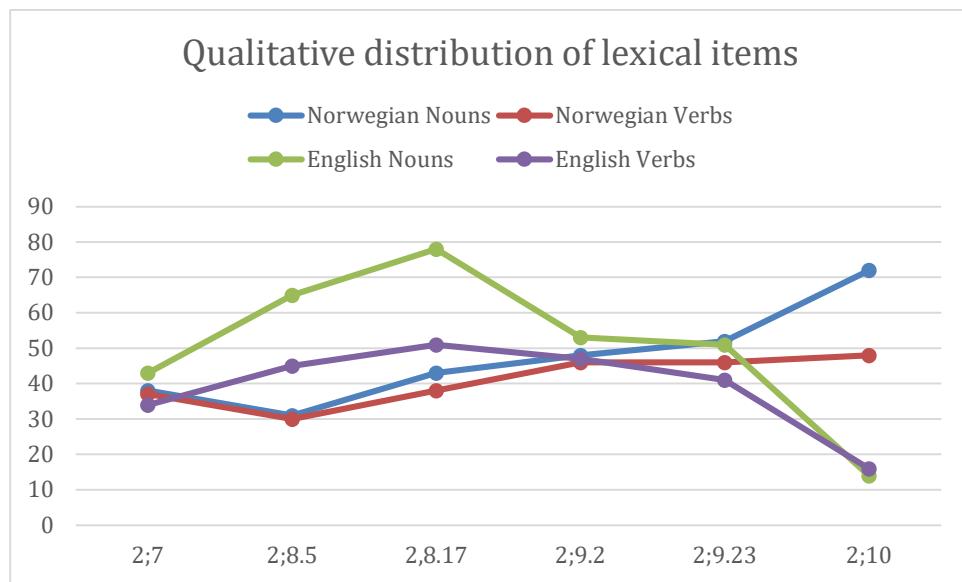
Across recordings her average use of different nouns is 50,7 for English nouns and 47,3 for Norwegian nouns. For verbs, the average use across recordings is 39,0 for English and 15,7 for Norwegian. This demonstrates again the prevalence of use of nouns over verbs for both languages, however, her use of Norwegian verbs is about half the average to English use of verbs. What is interesting when looking at the distribution across recordings is that although there are more nouns than verbs in her vocabulary in Norwegian, the relative distribution between nouns and verbs in her expressive language is small, except for the last recording, where there is a large difference of 24. At the most for the other recordings there are 8 nouns differing, but also two instances as low as only 1 more noun than verb used. Another interesting aspect of the distribution across recordings is the seemingly steady increase in her Norwegian. Whereas her English varies more.

4.1.3 Total and conceptual vocabulary

For Emma's total vocabulary number of different nouns and verbs across both languages were counted. This added to 317 nouns and 279 verbs, which then added to a total vocabulary of 597 words. There is high individual variety of vocabulary development and size (De Houwer, 2009) Because of this, it is difficult to establish an appropriate normative sample of the bilingual vocabulary, as bilingual experience varies to a great degree (Core et al, 2013). Therefore, a conclusion about these numbers is not easy to draw. However, compared with research that has previously looked into total vocabularies have also found bilinguals' total vocabularies around 30 months of age to be around 500 – 600 words (Core et al, 2013).

Table 6. *Numbers for total vocabulary.*

	English	Norwegian	
TDN	204	183	
TDV	114	96	
TV	318	279	**597**

TDN= Total number of Different Nouns, TDV= Total number of Different Verbs, TV= Total Vocabulary.

Emma's conceptual vocabulary was found by subtracting total number of overlapping nouns across both languages from total number of different nouns across both languages and subtracting total number of overlapping verbs across both languages from total number of different verbs across both languages. Adding these two numbers for nouns and verbs then made her conceptual vocabulary count, as seen in table 5. Adding her conceptual knowledge of nouns and verbs Emma displays a conceptual vocabulary of 463 words, compared to her total vocabulary of 597, as seen in figure 3 and 4. The result of her total vocabulary being larger than her conceptual vocabulary, as demonstrated in figure 3, is a result that matches previous studies where both similar and dissimilar language pairs have been studied and compared (Bosch & Ramon-Casas, 2014). Differences between total and conceptual vocabulary was expected, as this has been demonstrated in most studies using these measures (De Houwer, 2009, Bosch & Ramon-Casas, 2014). In both total and conceptual vocabularies, the noun prevalence is evident.

Table 7. *Numbers for conceptual vocabulary.*

	Nouns	Verbs	
TD	387	210	
Overlap	71	63	
CV	316	147	**463**

TD= Total number of different nouns/verbs across both languages, CV= conceptual vocabulary

Figure 3. Comparison of Emma's total and conceptual vocabulary.
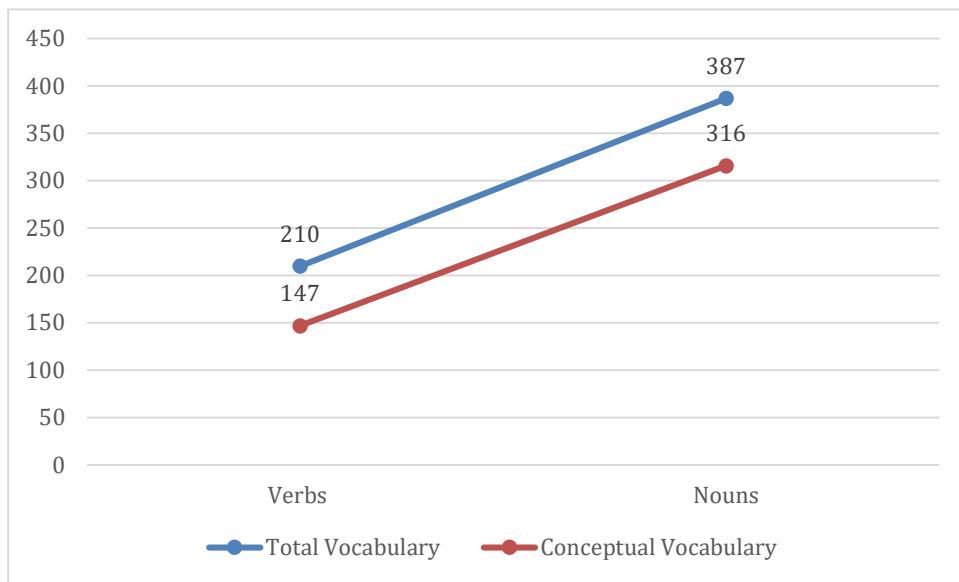
Figure 4. Comparison of Emma's total scores of total and conceptual vocabulary.
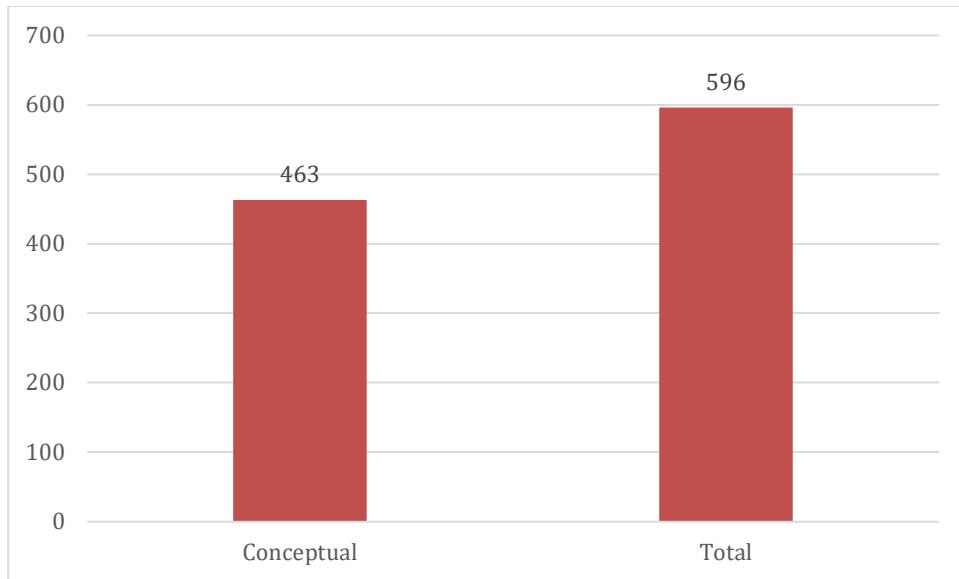
4.1.4 Sum-up

Emma displays relatively similar MLU scores in both her languages across recordings, which indicates that she is a relatively balanced bilingual. Her TTRs indicates a slightly more lexically varied language in her English, which could suggest that her English is slightly more advanced than her Norwegian. This is further supported through Emma's vocabulary scores, where she demonstrates a higher number of both nouns and verbs in her English than her Norwegian. In both her languages though, Emma ha a noun prevalence, something that is also evident in the lexical distribution across recordings. Across recordings Emma demonstrates a total vocabulary of 597 words, whereas her conceptual vocabulary makes up 463 words.

4.2 RQ II: Are there instances of translation equivalents in Emma's vocabularies, and if so, to what extent?

The comparison of Emma's vocabularies in terms of nouns and verbs in the two languages and the calculation of her conceptual vocabulary reveal many instances of translation equivalents (See appendix 2A and 2B for complete overview of her verb and noun overlap). Examples are her use of nouns like *train-tog, dress-kjole, hair-hår* and verbs like *hide-gjemme, see-se, wash-vaske, hear-høre.* When looking into Emma's translation equivalents, or doublets, there is a small difference in noun and verb overlap. With a total amount of noun overlap of 71 nouns and a verb overlap of 63 verbs, there is a small difference between the two lexical categories. However, Emma's verb overlap is much larger relative to total number of different verbs than her nouns overlap. When calculated from total number of different nouns, 18,3 % of her nouns are overlapping, whereas 30% of her verbs are overlapping when calculated from total number of different verbs. With an evident noun prevalence in her vocabularies, such a large verb overlap compared with noun overlap is interesting. With a total amount of overlap of 134 words, from both lexical categories, this makes out 29 %, from her conceptual vocabulary and 22,4 % from her total vocabulary. Compared with previous findings (see table 1, from Bosch & Ramon-Casas, 2014), her overlap falls within the 20 – 43% found in these studies, where the ages vary from 7 months to 3 years. As mentioned in the background section, language similarity could also influence overlap (Bosch & Ramon-Casas, 2014). To address this cognates and phonologically similar words were counted. This is demonstrated in table 5, where an overview of cognate words from her overlap is listed for both verbs and nouns. She has 37 cognates in her vocabularies, where 16 of them are verbs and 21 of them are nouns.

From Emma's overlap of 134 words, this makes 27,6 %. Here Emma's use of verbs such as *see-se, can-kan, sit-sitte* and nouns such as *finger-finger, cake-kake, balloons-ballong, sock-sokk* are a few of the cognates listed, illustrated in Table 5.

Table 8. Cognate nouns and verbs.

Cognate verbs		Cognate nouns			
Eng	Norw	Eng	Norw	Eng	Norw
like	like	hand	hånd	dinosaur	dinosaur
see	se	music	musikk	egg	egg
fall	falle	thing	ting	hat	hatt
pack	pakke	baby	baby	pizza	pizza
find	finne	ball	ball	sock	sokk
help	hjelpe	cake	kake	box	boks
set	sette	finger	finger		
sit	sitte	fish	fisk		
will	vil	foot	fot		
dance	danse	hammer	hammer		
fly	fly	man	mann		
ring	ringe	room	rom		
match	matche	teletubby	teletubby		
hold	holde	balloon	ballong		
hang	henge	book	bok		
put	putte	cat	katt		

Eng=English, Norw=Norwegian

4.2.1 Summary

Emma's vocabularies contain 134 TEs, which make out 29% f her conceptual vocabulary and 22,4% of her total vocabulary. She has a much larger overlap for action words, verbs, than she does for content words, here nouns. In addition, 27,6% of her overlap are cognates.

4.3 RQ III: Are there instances of code-switching in Emma's utterances and if so, what role do the code-switching instances play?

4.3.1 Code-switching in Norwegian

In Emma's Norwegian there are occurrences of code-switching. If one looks at table 6, it is, however, clear that these instances are few. With an average percent of 0,7% occurrence of Norwegian words in her English language use, ranging from 0 – 1,5 %, this does not

represent a large percent of code-switching instances. This is, however, a mere count of words occurring. The way in which these words are used, i.e., whether they are used as single items in a Norwegian sentence or whether they are used as whole sentences will be explored later. An interesting note though is that there are no single words that stand out to be used repeatedly. There are mostly words used that one time. At one recording there are even no occurrences of code-switching.

Table 9. English code-switched words.

Code-Switching in her Norwegian					
2;7.21	2;8.7	2;8.20	2;9.11	2;9.25	2;10.9
1 Piglet	1 grandma	3 I	1 a		1 a
1 look	1 grandpa	1 all	1 here		1 an
	1 in	1 clothes	1 lots		1 and
	1 office	2 come	1 money		2 black
	1 pink	2 daddy	1 of		1 calf
	1 shoes	1 goes	1 reindeer		1 need
	1 the	2 in	1 there		1 we
	1 trim	1 is			
	1 and	1 know			
		1 pink			
		1 not			
		2 popped			
		1 that			
		1 the			

4.3.2 Code-switching in English

In Emma's English there are several occurrences of Norwegian words. When comparing amount of code-switching in Norwegian and English in recordings one can see from table 7 that there are much higher percentages of Norwegian words in her English use than the other way around. Here the average occurrence of 6,2%, ranging from 2,6 – 13,9 %, of Norwegian words among her English words demonstrates a clear asymmetry in code-switching between the two languages. This is also in accordance with Bentzen's (2000) conclusions regarding Emma's code-switching directionality and asymmetry, where a clear asymmetry for Norwegian code-switching in Emma's English was found. Unlike the occurrence of English words in her Norwegian, there are a few words here that seem to occur repeatedly. Her use of *du (*you), occur multiple times at almost every recording. In three of her recordings the use of the word you occur over 20 times. This has, however, been discussed by the investigator (Bentzen, 2000)

to be a case of borrowing, as it has been hard from the recordings to make out a clear distinction between *du* and *you*, when it comes to pronunciation. In addition, it has also been pointed out by investigator that some of the Norwegian words are also only learned Norwegian. That is, though her parents both use English at home, some of the words are presented to Emma in Norwegian, the words *barnehage* (kindergarten) and *sånn* (there/like that) are also used in Norwegian by her parents (Bentzen, 2000). The use of these words in her English will be further addressed in the next section.

Table 10. *Norwegian code-switched words.*

Code-Switching in her English					
2;7.14	2;8.5	2;8.20	2;9.2	2;9.23	2;10.8
1 barne	1 brunost	1 andre	2 den	1 å	3 æ
1 barnehage	1 dem	2 den	21 du	1 æ	3 Brumm
1 den	2 den	1 derre	8 er	1 avdeling	6 du
8 du	1 det	1 det	1 hør	5 barnehage	2 falt
1 har	28 du	1 dokker	6 hadet	1 biola	3 Tigergutten
1 ikke	1 går	27 du	4 hallo	1 dør	2 ho
1 ka	7 hår	1 en	1 har	1 det	2 ja
1 kommer	1 har	7 er	1 hei	14 du	3 jo
1 og	1 i	1 hei	6 heia	2 dukken	3 leka
1 sånn	1 ikke	11 heia	7 hjemme	2 en	3 med
1 saft	1 ja	5 hjemme	8 ja	1 gå	3 og
1 se	14 sånn	1 må	1 kor	1 glass	1 sånn
	2 sånns	1 på	2 litt	3 hør	2 tippa
	1 sa	1 morra	1 så	1 har	
	1 vegg	3 sånn	2 sånn	1 kan	
		1 ta	1 sint	2 litt	
		1 var		1 må	
		7 hadet		5 matpakke	
		1 men		1 legge	
		1 her		1 så	
				1 sånn	
				1 sår	
				1 sæ	
				1 tøm	
				1 toget	
				1 vaske	
				1 farfar	
				1 farmor	
				1 frosk	

4.3.3 Word types in Emma's code-switching

Table 6 and 7 further give an overview of not only which words and how many times they occur, but also how many function words and how many content words. By content words here is meant all the words that contain a meaning in themselves, like nouns, verbs, adjectives, and adverbs, whereas function words entail words that contain little meaning on their own and rather create structural relationships for the content words, such as pronouns, prepositions, determiners, and conjunctions. From this table it is evident that her majority of code-switching words are contentives. This would suggest a lexically supportive function of her code-switching (Lanza, 1992). Something that would support the CP implications of use of code-switching, which will be further elaborated on in the discussion.

4.3.4 The role of her code-switching

4.3.4.1 The role of code-switching in her Norwegian

When looking closer at her use of English in her Norwegian language, two different patterns stand out. If she code-switches in entire sentences this almost exclusively occurs when she is speaking with her mother, as seen from (3) where N and E stand for Norwegian and English respectively:

> (3)   **I not know**. (talking to her mum)   2;8.20
>   (N: jeg vet ikke)
>   ja.
>   (E: yes)
>   den var ikke igjen der.
>   (E: it was not left there / it was not there again)
>   æ skal bare ordne den her.
>   (E: I am just going to fix this (one))
>   ja, jeg har.
>   (E: yes, I have)
>   nei.
>   (E: no)
>   **I like to daddy come in**. (talking to her mum)

(N: Jeg vil at pappa skal komme inn)

I like to daddy come in.

(N: Jeg vil at pappa skal komme inn)

Here she very clearly distinguishes between interlocutors, where she switches to full English sentences when speaking with her mother, with whom she is usually speaks English. She then switches back to Norwegian when going back to playing with the investigator.

The other pattern that stands out in her use of code-switching in Norwegian, are English words occurring in her Norwegian sentences. Like in (4), (5) and (6):

(4)   Jeg skal hente dem.   2;8.7

(E: I will get them.)

 Det er <**trim shoes**>.

(E: They are trim shoes)

Jeg skal hente dem.

(E: I will get them)

Du må hjelpe å finne andre.

(E: You must help find the other (one))

Jeg finn bare en joggesko.

(E: I can only find one trim shoe)

(5)   Og en rød ballong her.   2;8.20

(E: And a red balloon here)

Den her andre ballong <**popped**>

(E: The other balloon (here) popped)

Den andre ballong <**popped**>

(E: The othe balloon popped)

(6)   Æ vet ikke.   2;9.11

(E: I don't know)

<**A reindeer**> ?

(N: Et reinsdyr?)

Ka det der er for nokke?

(E: What is that?)

.

In the case of (4), with the trim shoes, she does not seem to remember the Norwegian word first, but then later on she uses it. This demonstrates that she knows the word, however, the English word seem to have been the word readily at hand. Which also seems to be the case for (5) and (6), where her English words seems to be more readily at hand than the Norwegian word, but where we cannot, from the recordings alone, know whether she knows the Norwegian word for this. Her use of code-switching in these instances thus seems to either be instances of changing communicative interlocutors or seem to have a supportive role as a communication strategy when the correct word is not readily at hand.

4.3.4.2 The role of code-switching in her English

When looking at instances of code-switching in her English, there are instances of both between utterances code switching as well as within utterances code-switching. There are, however, several more instances of Norwegian words in her English language, although several of those are with the same words, *du* and *sånn* (you and there/like this). Here, as mentioned, there are issues regarding the ease with which *you* and *du* could be transcribed, which could in turn make it difficult to draw any conclusions regarding her code-switching between these two words. However, from the recording where she uses the most instances of *du*, she also uses at least two clear instances of *you*. Below, the question marks are kept from the transcripts to indicate instances where it is not certain whether she uses the Norwegian or English word.

    (7)   No. Go out!   2;8.5

            (N: Nei. Gå ut!)

            Du (?) need to find.

            (N: Du må finne)

            I can hide you.

            (N: Jeg kan gjemme deg)

……

(8)   Those trucks gonna drive.   2;8.5
             (N: De lastebilene skal kjøre)
             Du (?) need to drive this truck, (du) (?) Daddy.
             (N: Du må kjøre denne lastebilen, du pappa)
             You need to drive that.
             (N: Du må kjøre den)
             I gonna drive ...
             (N: Jeg skal kjøre…)
             Du (?) need to drive den her (?) truck.
             (N: Du må kjøre denne lastebilen)
             Yellow
             (N: Gul)
             and this big truck is so big!
             (N: Og denne lastebilen er så stor!)
             this truck is so big.
             (N: Denne lastebilen er så stor)

From this excerpt of her conversation during this recording we can see that in both example (7) and (8) she clearly uses *you*, as well as instances of *du*. In addition, in example (8) there is also an instance of her using *den her* instead of *this,* although she uses *this* later in the conversation. These are both instances of using the Norwegian word first, and then changing it to the English word later, as is also seen in her English code-switched words when speaking Norwegian. This could indicate lexical gap filling.

What is also a relevant aspect of code-switching is the language mode called for by the context. As the CP predicts, Emma's background calls for a monolingual language mode. This is also demonstrated as her mother reacts to her use of Norwegian when Emma is trying to call herself on the phone. Here MOT is indicative of the mother speaking, whereas EM here indicates Emma as speaker.

(9)   MOT: Emma, can you call your friend Emma?   2;8.17
             (N: Emma, kan du ringe vennen din Emma?)

EM: Yes.

(N: Ja)

EM: Heia Emma, er du ...

(E: Hey Emma, are you…)

MOT: no, she only speaks English.

(N: Nei, hun snakker bare Engelsk)

EM: huh?

MOT: remember Emma only speaks English.

(N: Husk at Emma bare snakker Engelsk)

MOT: you have to speak English.

(N: Du må snakke Engelsk)

EM: Heia, Emma, is du

(E: Hey Emma, is you)

EM: is du (?) home, Emma?

(N: er du hjemme Emma?)

Here her mother is reminding her that she must speak English. This demonstrates an explicit expectation of which language Emma should use, which calls for a monolingual language mode. However, what precedes and follows this excerpt is also relevant. Before her mother asks her if she can call Emma, Emma has been playing that she calls her friends, to whom she speaks only Norwegian, where her sentences are only Norwegian as well. So, she has made the phone calls in Norwegian. As soon as she talks to herself, she starts speaking Norwegian, but then her mother reminds her that she must speak English. Helping her back to her monolingual language mode. What follows when she then calls one of her friends again, however, is this:

(10)   That ringed.   2;8.17

            (N: Det ringte)

            Heia (.) heia.

            (E: Hey)

            Maybe that is [/] maybe that is Angus?

            (N: Kanskje det er/ kanskje det er Angus?)

            Heia, Angus!

            (E: Hey Angus!)

Er du home, Angus?

(E: Are you home, Angus?)

Ja bye bye, Angus?

(E: Yes bye bye Angus?)

Ja.

(E: Yes)

Now have I ringed Angus.

(N: Nå har jeg ringt Angus)

Heia (.) heia Adrian!

(E: Hey. Hey Adrian!)

Is dokker hjemme, ja.

(E: Are you at home, yes.)

Hadet.

(E: Bye)

From speaking with her friends in only Norwegian sentences, she has now started to code-switch after she spoke only English when calling herself. Now she is bringing English words into her Norwegian conversation with her friends.

An interesting instance of code-switching is where she seems to start code-switching to Norwegian when she is playing with her dolls:

(11)   I washing it.   2;9.23

(N: Jeg vasker den)

æ vaske dukken så dukken kan gå å legge sæ.

(E: I am washing so the doll can go to bed)

COM: I think she is speaking Norwegian to her dolls here...!

(N: Jeg tror hun snakker norsk til dukkene sine her...)

this is all gone.

(N: dette er alt borte)

......

As she is in an English monolingual language mode in her own home, where she speaks English with her parents, this switch does not seem to be based on contextual cues by her

interlocutor. This might suggest that there could be other instances, other than what the context "requires" that influences instances of code-switching. Could this be a game she frequently plays in her kindergarten that she brought home with her? However, conclusions, other than her use of more code-switching when talking in English rather than Norwegian, can only be speculative.

4.3.5 Summary

Emma demonstrates instances of code switching in her English and her Norwegian, where there are more instances of code-switching in her English than her Norwegian. Most of these code-switching instances, in both languages, are content words.

There were two patterns in Emma's code-switching in Norwegian. One where she switches between interlocutors, where she switches from Norwegian with the investigator to English with her mother. Here she switches from full Norwegian sentences to full English sentences. The other pattern is where she uses an English word midsentence, where she later uses the correct Norwegian corresponding word.

In Emma's English the code-switching also displays the same patterns, where she switches to full Norwegian sentences when calling her friends, whereas otherwise she speaks full English sentences. She also demonstrates several instances of using Norwegian words midsentence, before later using the corresponding English word.

Example (9) demonstrates how Emma's parents actively preserves a monolingual language mode for Emma, depending on who she communicates with. (10) and (11) demonstrates further instances of code-switching where it is suggested other factors may influence.

4.4 RQ IV: Do the characteristics reflect the differences in Emma's language exposure and language use patterns? Are her vocabularies distributed (home vs. daycare)?

As seen, Emma's characteristics demonstrate that she is, according to her MLUs and TTRs, a relatively balanced bilingual. Where her English language might be slightly more advanced, as it is more lexically diverse and with a slightly larger vocabulary. Though, her

MLUs indicate that her languages are fairly balanced in terms of sentence complexity, her TTRs indicate a more lexical diversity in her English language, which was further supported through her larger number of English nouns and verbs compared to Norwegian. However, these differences are not excessively big, indicating, as previously mentioned, that she is a rather balanced bilingual. In addition, the fact that her code-switching utterances are a part her MLU score, might also influence this outcome. These results would seem to reflect her exposure background well, as she has most likely spent her first year at home developing her English, before she started attending kindergarten and expanding her exposure to Norwegian. Which could explain a richer vocabulary in English than Norwegian, as her Norwegian is in a period of expansion. The languages are similar, with similar word classes, a basic SVO sentence structure, and several words that are similar and/or identical. In addition, both languages emphasize an object saliency, which would also accommodate the noun prevalence then displayed in both her languages. Here, however, one could ask whether her high number of verbs compared to nouns in her Norwegian language use pattern could possibly be due to the necessity of action words in her daily activities, as she has most likely learned some of these in English already and whether these could have facilitated acquisition in Norwegian. Here research demonstrate that for language pairs with similar words, TE's that share form similarity are more easily acquired, the question then would be whether her knowledge of action words in English has a facilitative role on her development of action words in Norwegian. This could potentially explain why her overlap consists of 63 verbs compared to 67 nouns, which is not large difference. That she then has as complex a sentence structure, even if her vocabulary might be slightly smaller, is also not surprising. This would again entail that her language characteristics reflect her language exposure patterns.

As far as her language use pattern goes, this has a surprising twist. Her amount overlap demonstrate that she uses a lot of the same words for her home domain and her kindergarten domain, which would indicate that her vocabularies are not necessarily distributed between home and day-care. However, as previously mentioned, activities in both contexts are not necessarily that different, which would entail a need for the same words across both domains. In which case, the CP clearly predicts a relatively large amount of overlap between her vocabularies that is also evident in her language use pattern.

What is surprising in her language use pattern, is her amount of code-switching. In both languages she demonstrates a clear division between languages based on interlocutors, something that supports the CP, where context plays a large role in her language choice. The

clear division between languages is compatible with her being in monolingual language modes and her switch between English with her parents and Norwegian with her investigator demonstrates a clear knowledge of the context "expectation" with regard to language choice. This sensitivity to contextual cues to language choice has been demonstrated in children as young as 1;9 years of age (Nicoladis & Genesee, 1997). Also, as previously mentioned, the CP predicts that her code-switching will be minimal due to the strict separation of languages according to interlocutor, emphasizing that instances of code-switching will be of supportive nature in communication. Here the asymmetrical pattern in itself is not necessarily surprising as several research studies has demonstrated more use of code-switching from the dominant language to support communication in the weaker language (Nicoladis & Genesee, 1997). What is surprising is the directionality of the transfer. Her MLUs, TTRs and her lexical item quantities indicated a balance between the languages, with perhaps a slight advantage to her English as her possible dominant language, if not at least the more advanced language. On the other hand, her code-switching pattern seem to indicate her Norwegian to be the stronger language, due to a larger amount of code-switching in her English. Here one could speculate whether this too could indicate as shift in her language acquisition. Although here one also begs the question of what her MLUs and TTRs could have been if code-switching utterances had been excluded.

As previously mentioned, her overlap between the languages falls within the 20 – 40% found in the overview by Nicoladis & Genesee (1997), however, her overlap makes out 21 % from her total vocabulary and 28 % from her conceptual vocabulary. Out of total number of different verbs used, 30% of them overlap between the languages. This is after some have been excluded due to isomorphic interpretation difficulties, which means she could have an even larger overlap between her vocabularies for activities she does at home and in her kindergarten.

4.5 RQ IV: How do Emma's expressive lexical skills compare to that of monolingual peers in each language?

4.5.1 MLU in English and Norwegian: Bilingual Emma vs. monolingual Shem and Ole

In English, I compare Emma to a monolingual child of the same age called Shem. When looking at Emma's MLUs next to Shem's MLUs, see table 8, here too the MLU's are of similar number. This indicates that her expressive language skill is not much lower than that of Shem's. There is a marked difference at one point, at age 2;9.1 for Shem and 2;9.2 for Emma, where

Shem's MLU is much lower, however, this could again be a day where he did not talk as much. As the recordings are done at close points in time, a developmental trajectory is not to conclude. Here too, the average MLU across the recordings indicate, with her 4,35 and his 4,23, that their expressive language skill are at the same level.

Table 11. MLUs Emma and Shem.

Ages (Emma & Shem)	Emma	Shem
2;7.14 & 2;7-10	4,06	4,41
2;8.5 & 2;8.3	4,53	4,1
2;8.17 & 2;8.20	4,15	4,1
2;9.2 & 2;9.1	4,31	3,67
2;9.23 & 2;9.27	4,65	4,42
2;10.8 & 2;10.2	4,37	4,66

In Norwegian, I compare Emma to a monolingual child of the same age called Ole. Comparison of MLU's between Emma and Ole shows a similar level of language development, with small variations in MLU across the different recordings, see table 9. Interesting here is that same peak in MLU at the age 2;8.20 for Emma and 2;8.24 for Ole, where both have a high score for MLU. The average for the MLUs across recordings is here too of relative similar level 4,17 for Emma and 3,91 for Ole, a difference of 0.26.

Table 12. MLUs Emma and Ole.

Ages (Emma & Ole)	Emma	Ole
2;7.21 & 2;7.20	3,72	4,02
2;8.7 & 2;8.5	3,94	3,85
2;8.20 & 2;8.24	4,94	4,83
2;9.11 & 2;9.15	3,92	3,34
2;9.25 & 2;10.0	4,13	3,62
2;10.9 & 2;11.23	4,39	3,85

Emma's expressive language skills thus seem comparable to those of her monolingual peers, where she produces the same length of her utterances as the monolingual peers do in their respective languages. This is in accordance with previous research on bilinguals, where they have been demonstrated to follow the same developmental trajectory as monolinguals (De Houwer, 2009).

4.5.2 TTRs in English and Norwegian: Bilingual Emma vs. monolingual Shem and Ole

Comparing lexical diversity could also shed some light on another aspect of vocabulary differences. In English, I find that TTRs for Emma range from 0,156 – 0,371, whereas Shem's range from 0,176 – 0,296. Here the average TTR for Emma is 0,244, whilst Shem's is 0,236, indicating that they both seem to be at the same level of lexical diversity and have equally advanced language. For an overview of the TTRs see Appendices 1A for Emma and 3A for Shem.

In Norwegian, Emma's TTRs range from 0,133 – 0,232, whereas Ole's TTRs range from 0,163 – 0,246. The average for Emma is of 0,179 and for Ole it is 0,188. Similarly, to what I found in English, the TTRs in Norwegian are rather similar for Emma and Ole, possibly indicating that their utterances are equally lexically diverse. For an overview of the TTRs see Appendices 1B for Emma and 3B for Ole.

4.5.3 Comparison of vocabulary scores

When looking at the separate comparison scores for her separate language vocabulary scores, here measured by total amount of different nouns and total amount of different verbs in each language, we can see from figure 3 that Emma scores below that of her monolingual peers in both respects. With 183 different nouns in her Norwegian, it is decisively less than Ole's 347 different nouns. The same is evident in English, where her 204 nouns are less than Shem's 318 nouns, though the gap is slightly smaller than that in her Norwegian. These results are, however, not unexpected, as several studies have demonstrated this gap to exist when bilingual's languages are measured and/or compared separately (Unsworth, 2013). If we look at both her total vocabulary score (TVS) and her total conceptual vocabulary (TCV) also in figure 3, the numbers are different and more comparable. Here her TVS is actually higher than her monolingual peers, with 387 nouns and 210 verbs, compared to Ole's 347 nouns, 175 verbs and Shem's 318 nouns and 146 verbs. Taking these numbers together in an overall vocabulary count, Emma displays a total vocabulary of 597 words, compared with Ole's 522 words and Shem's 464 words. This has also been demonstrated in other studies, where bilinguals match or even outperform monolinguals when both their languages are taken into account (Bialystok et al, 2010, Unsworth, 2013). This demonstrates that her total vocabulary is larger than those

of her monolingual peers, it is just distributed across two languages, rather than one.

Figure 5. Lexical item quantity/vocabulary scores.
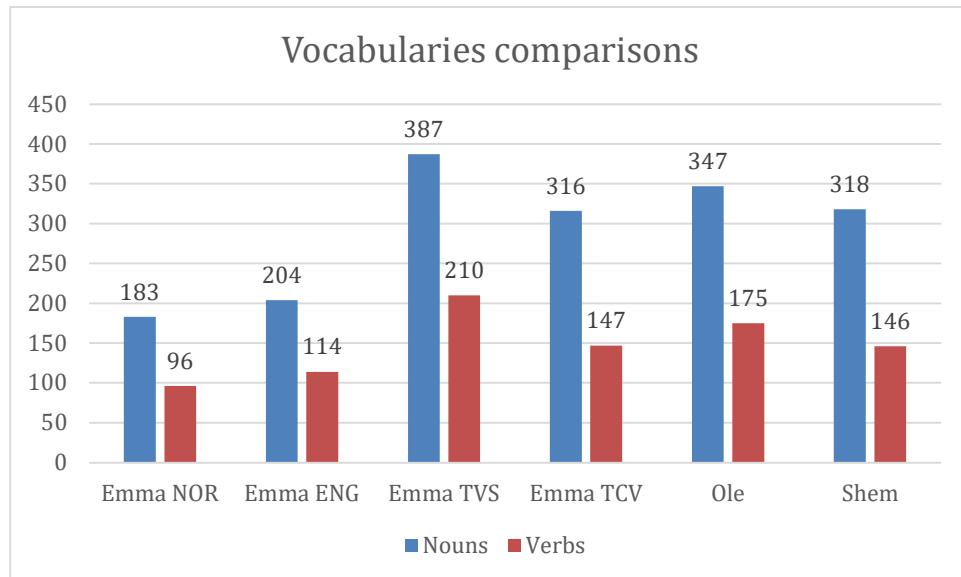
Emma NOR= Emma's Norwegian, Emma ENG= Emma's English, Emma TVS= Emma's total vocabulary, Emma TCV= Emma's conceptual vocabulary.

This is not the quite same, however, for her TCV. Here her 316 nouns and 147 verbs, do not quite match up to the numbers of Ole. However, her TCV compares with Shem's number of nouns and verbs. This demonstrates that her TCV score is comparable, at least to that of the English monolingual. However, as demonstrated by (Schelletter, 2002, Bosch & Ramon-Casas, 2014), the similarity of languages matters too in the development of amount of TE's, where similar languages, with many similar word-forms could accommodate an increased amount of TE's. This would would then cause the TCV to become smaller as only one of the words for the same concept is being measured. Here English and Norwegian are similar languages, with several form-similar words, which could indicate that her lower TCV compared to the vocabulary level of Ole, could be due to her relatively large amount of overlap.

4.5.4 Summary

Both Emma's MLUs across recordings and her TTRs across recordings were comparable to Shem's and Ole's. This indicates that she exhibits the same level of complexity and diversity in her English and Norwegian language as that of monolingual peers.

Comparing Emma's vocabularies to those of Shem and Ole demonstrated similarities and differences. They all exhibited a noun prevalence in their vocabularies, see figure 5. Emma's separate vocabulary for each language was compared to those of her monolingual peers, where her vocabularies for the separate languages were much smaller than those of Shem and Ole. Something that is consistent with previous research (Unsworth, 2013). Emma's total vocabulary of 597, was larger than both Ole's vocabulary of 522 and Shem's vocabulary of 464. Additionally, Emma's conceptual vocabulary of 463 was comparable, almost identical to Shem's monolingual vocabulary of 464. Though it was smaller than Ole's vocabulary of 522.

5 Discussion

5.1 What are the characteristics of Emma's expressive language skills in the heritage language, English, and the majority language, Norwegian?

5.1.1 Emma's MLUs and TTRs

Emma's average MLU scores were 4,35 in English and 4,17 in Norwegian and her average TTR scores were 0,244 in English and 0,179 in Norwegian. A comparison in TTRs and MLUs indicate then that Emma is a relatively balanced bilingual, with a language use pattern indicating that her vocabulary might be more diverse in English than Norwegian and a possible slight indication of dominance in English. A question is though whether her MLUs might have demonstrated a different relationship between the languages if code-switching utterances had been left out of the MLU analyses. Other research has demonstrated how some of the longer MLU utterances, were utterances involving code-switching (Quick et al, 2018). This could influence the resulting MLU values of Emma. However, in this case they were not excluded, and results are discussed accordingly. Emma's results seem to contrast the prediction from the CP's implications, that her English might be less developed and less dominant than her Norwegian, as it is limited to fewer people and fewer domains, i.e., her parents and her home. In contrast, she speaks Norwegian during large proportions of the day at her kindergarten and with more people. While the input in Norwegian may be larger than in English for Emma at the time of data collection, Emma's overall exposure to the two languages from birth may be rather similar, since she has spent her first year of life at home with her parents, who both speak English. Therefore, her amount of English exposure might not be any less than her amount of Norwegian exposure so far, which may explain the fact that her MLUs and TTRs in the two

languages are rather similar. As demonstrated by Thordardottir (2019), amount and age of exposure are both relevant factors in language acquisition, where amount seems to have an advantage over age of exposure.

An interesting question here would be whether her balance between the two languages results from an increase of majority language exposure during her second year of life when Emma started the kindergarten. It can be hypothesized that the input in Norwegian will increase in the following years and may lead to a Norwegian language dominance eventually. If her expressive language skills would change over time then, as her exposure to Norwegian might outweigh her exposure to English, this would be in accordance with the implications from the CP on how dominance can change over time across domains and situations. This is of course a speculation.

5.1.2 Lexical items quantity and distribution

Emma demonstrates a noun prevalence over verbs both in terms of number of nouns in her vocabulary and distribution of noun use across recordings. This seems to follow the monolingual trend for many languages (Childers & Tomasello, 2006). Something that is explained through objects being more salient in their surroundings and thus more easily mapped onto objects, whereas verbs might be less salient and in addition require some level of abstraction and generalizability to be learned independently from context (Gentner, 2006). Research has also found, however, that this does not need to be the case, where studies have found how quality of input, like parental speech, also can influence number of nouns and verbs in vocabulary development (Qiu & Winsler, 2017). For Emma's result this does seem to follow the noun prevalence found in many other studies. In addition, this also meets the prediction in terms of Norwegian and English being similar languages, both regarded as object salient languages (Gentner, 2006). The number of nouns and verbs across the languages of Emma also demonstrates what her TTRs indicated: she knows somewhat more words, both nouns and verbs, in English than in Norwegian with 204 nouns in her English and 183 nouns in Norwegian, and 114 verbs in English and 96 verbs in Norwegian. Her MLUs demonstrated her utterances to be of same length for both Norwegian and English. Her Norwegian utterances, however, are not as varied as her English utterances, according to her TTRs. This was explored in distribution of nouns and verbs in her speech across recordings.

The distribution of nouns and verbs across the recordings demonstrated that when the total number of nouns and verbs used was compared, Emma used more verbs than nouns in her speech. However, as the same verb can be used repeatedly for different objects, total number of different nouns and verbs used was compared. Here Emma demonstrated a larger use of different nouns over different verbs. Which is also in correspondence with her TTRs. What was interesting about her results here though, was how the gap between the number of different nouns used compared to different verbs used in her Norwegian was small. Even though her vocabulary indicates a larger portion of nouns over verbs, her Norwegian use demonstrates a larger use of verbs relative to nouns than does her English use. In addition, across recordings her Norwegian demonstrates a seemingly steady increase (see figure 2). This could again be due to verbs being more universal across situations then is perhaps the objects Emma encounters across domains, or reflect that Emma has a larger noun vocabulary in English than Norwegian. This could indicate that there is a slight dominance to her heritage language over her majority language. According to the CP, again, her heritage language would be predicted to be less dominant as it is restricted to fewer domains with less people with whom to use her language. Here the CP emphasize the ability to use a language is paramount for its development (Grosjean, 2015). However, as mentioned, she has most likely spent her first year at home, speaking mostly English, whereas increasing exposure to Norwegian through kindergarten could indicate that her Norwegian is in development, and that is why her language might be more lexically diverse in English and her vocabulary slightly larger, whereas her Norwegian seemingly demonstrates a steady increase.

A conclusion can of course not be drawn from a relatively limited time span such as this and where speech production may vary according to mood and activity. However, an intriguing question is whether this might indicate a shift in her language dominance. From a perspective of the CP, attendance to kindergarten might lead to increased amounts of exposure to Norwegian, whereas her English will remain in the same domain, with the same people. A shift in dominance is thus, according to the CP not unlikely, as it emphasizes how languages, and their fluency and dominance are in continual changes (Grosjean, 2015). This is also in line with the study from Montanari et al. (2018) who demonstrated how children's vocabularies in their majority language changed during the first 4 years of school. In their study the vocabulary in the children's heritage language stagnated, while their vocabulary in their majority increased. However, as Emma is in a strongly developmental age, the

difference seen in her languages could also merely reflect a differential leap in her languages that later will even out, as research has demonstrated that the vocabularies of bilingual children do not necessarily develop at the same rate (De Houwer, 2009). These are of course speculations, a longitudinal study, or cross-sectional studies at several points in time are needed to look further into this matter. For now, Emma's lexical item distribution shows, along with the quantity of nouns and verbs within and across languages, that her English is slightly more advanced and the distribution of nouns and verbs for both languages follows what has also been demonstrated in previous research, namely the noun prevalence (David and Wei, 2008, De Houwer, 2009).

5.1.3 Total and conceptual vocabulary

Regarding the total and conceptual vocabulary, Emma demonstrates a larger total vocabulary (597 words) than conceptual vocabulary 459 (words). This difference is not unexpected, as total vocabulary includes all different nouns and verbs in both languages, whereas conceptual vocabulary is a count of all concepts known. A concept for which she knows a word in both languages is counted only once, meaning some of her words are not included in the count. How these results meet the predictions from the CP is difficult to conclude, as it depends on how one looks at her domains in this respect. On the one hand, if the home and kindergarten domains are viewed as separate domains, the results do meet the predictions from the CP that her total vocabulary should be larger than her conceptual vocabulary as it encompasses words known from the home and kindergarten domains. On the other hand, if the domains are looked at as separate domains, this should entail a small overlap, which would also entail a large conceptual vocabulary. This is partially demonstrated. Compared with the total vocabulary of 597 words, her conceptual vocabulary of 459 is not a small conceptual vocabulary. According to location and interlocutors, the domains are separate and require the use of different languages. However, the activities in both domains might not be very different. Therefore, to which extent the domains are shared or separate is not easily concluded. This will be further discussed in the following research questions.

However, here there were methodological issues. One of the issues raised with counting conceptual vocabulary is how this measure of vocabulary assumes an isomorphic equivalent between concepts, and for languages such as Norwegian and English, there are a lot of the nouns that are, such as her use of *dress-kjole, doll-dukke, hide-gjemme*. In addition, her

vocabulary contains cognates, such as the words *dinosaur, baby,* form-similar words such as *balloon-ballong*, *mommy-mamma, sock-sokk, cake-kake* and form-dissimilar words like *rabbit-kanin, dress-kjole*. All these words are isomorphic translations. There are, however, also words here that are not isomorphic, here she uses both the noun *chick* and the noun *chicken,* which have two different semantic meanings in English, whereas in Norwegian there is only *kylling*, and this is used for both. This also goes for her use of the words *bunny* and *rabbit,* both of which would be called *kanin* in Norwegian. This became particularly difficult when counting verb overlap. Different pragmatic uses entail that the same verb in English can mean different things depending on situation, for example, when encountering verbs as *walk* and *go* in Emma's English, which should be the translation equivalent to her Norwegian *gå*. This was also the case for several other verbs, particularly some of her auxiliary verbs. Counting only one of these possible translation equivalents then entails that not all of Emma's words could be taken into account. This is an issue that has been raised before in relation to conceptual vocabularies (Core et al, 2013). An additional issue is how counting concepts known, entails that a word from either language that is also known will not be taken into account, and this has been emphasized to undermine lexical knowledge in bilinguals. This is because there are phonological processes behind learning words and as such these words learned are part of a bilingual child's lexical skill (Core et al, 2013). For Emma, some of her verbs were not included that possibly could have given a clearer image of her lexical items and thus her language characteristics.

The characteristics of Emma's expressive language skills so far then indicate that she follows the same developmental pattern found in previous research (Lanza, 1992, Gentner, 2006), with a noun prevalence and dominance of noun use in her language use pattern. This could also seem to concur with the predictions made by the CP. If one follows Gentner's (2006) argument of object saliency and its relation to prevalent and salient object use in different languages, her exposure to the languages in separate domains, along a monolingual language continuum mode, could be said to further a contextually based noun prevalence. Thus, expressive language skills reflection of contextual influence would concur with the premise from the CP.

As for Emma's total and conceptual vocabulary levels, the CP does emphasize the need to include both languages when looking into vocabulary knowledge, as different contexts will allow for development of different vocabularies. To gain a clear picture of the bilingual's total

vocabulary knowledge, differing contexts possibly creating differing sets of vocabularies, must be taken into account. As young simultaneous bilingual vocabularies have not been extensively studied in light of the CP, amount, or totality, of expected vocabulary has not been established. Though one could argue that as the CP emphasizes context and contextual influences, and not individual, cognitive, influences, an amount of vocabulary might not be an area the CP is "responsible" for. As Grosjean (2015) states, the CP should be seen in connection with other phenomena characterizing a person's bilingualism.

5.2 Are there instances of translation equivalents in Emma's vocabularies, and if so, to what extent?

There were several instances of TEs in Emma's vocabularies, where 67 nouns were found to overlap, and 63 verbs were overlapping. This made out 21% of her total vocabulary and 28% of her conceptual vocabulary. This is not perhaps a very large overlap, which could indicate relatively distributed vocabularies, as some basic overlap between vocabularies might be expected (Grosjean, 2015). What could point toward whether her vocabularies are distributed or not, however, is the large amount of overlap for action words. The slight difference between nouns and verbs is interesting as her English and Norwegian vocabulary both have larger amounts of nouns than verbs. But here 18,3% of different nouns used and 30% of the different verbs used were overlapping. Action words might be more universal sometimes, than objects used, and therefore create a larger overlap for verbs than for nouns. The verbs like *play*, *dance* and *climb* are universal actions, whereas what she plays with, where she climbs, might differ between domains. It could also be due to her vocabulary at this age being more universal and basic, as several of these verbs are basic words that will be needed across several situations and domains. The fact that she knows several of them in both English and Norwegian suggests she uses them in both her domains and therefore these parts of her vocabularies are not distributed, but shared. Large portions of her vocabularies, however, are not shared, and therefore her vocabularies as a whole seem to be distributed across domains, as predicted by the CP.

These numbers indicate that her amount of translation equivalents might not constitute a very large overlap, but relatively large. Though looking at the lexical categories separately, would suggest that she has a larger overlap for verbs, making out 30 % of her total number of

different verbs compared with the 18,3% of nouns from her total number of different nouns. This could indicate the universality of actions over objects, as mentioned. However, another reason for her relatively large amount of overlap could also be due to language similarity. As Bosch & Ramon-Casas (2014) demonstrated, language similarity could influence acquisition through presence of phonologically form-similar words. Phonologically form-similar words, here especially cognates, facilitate early lexical acquisition. English and Norwegian share several cognates and form-similar words, something that could also influence amount of overlap. Thus, underlying factors for her overlap might be due to both language pair in question as well as contextual factors. This highlights another challenge when investigating bilingual vocabulary acquisition. Due to individual variety, in terms of language pairs and contextual background, what is comparable across differing bilingual linguistic backgrounds is a challenge. And as is emphasized by the CP, an overlap would more be a condition of contextual background, i.e.., language use patterns connected to domains and interlocutors. However, as seen above, this is difficult to separate from other influential factors.

Here again, methodological issues must be addressed. As there is a considerable number of verbs that are not isomorphic, the count of verb overlap is especially problematic. Here there are words that have been excluded due to multiple meanings, as only one word has been counted as a translation equivalent. Thus, her overlap could be larger than expressed here. However, by comparing the noun overlap alone, not only does the amount of translation equivalents look smaller, one loses an important aspect of her vocabulary as well. There are several verbs that are not isomorphic and hence difficult to account for as a TE, although she clearly can express these instances pragmatically, such as the Norwegian *må*, would translate to her use of English *have to.* But if *have* is already an equivalent of the Norwegian *å ha*, can it also be counted as an equivalent of *må?* As previously mentioned, only one of the equivalents have been counted in order to be concise in the counting of TEs. Therefore, the most direct, and most isomorphic translation equivalents have been connected. However, and again as previously mentioned, this also entails that not all TEs are accounted for. Verbs are still accounted for as a comparable category because there is a number of TEs that are not only isomorphic, but also very similar. Words like *dance-danse, fall-falle, can-kan, see-se*, are instances of TEs in Emma's vocabulary that are almost identical, and which can have an influence on her vocabulary.

Demonstrating a relatively large amount of overlap between her vocabularies, it is difficult to draw any conclusions regarding the implications of the CP. First, the fact that Emma has a relatively large overlap does not support the implications of the CP as it suggests that her

domains are not shared. Although it must be emphasized that activities might be of similar nature, which could indicate her domains to be shared. This would explain her overlap. Here it is questioned whether her verb overlap result is due to the basicness of her vocabulary, or whether the universality of action words might be why a large portion of her verbs are shared; the rest of her vocabulary to be distributed, however. The fact that Norwegian and English are similar languages and share several cognates and phonologically similar words, could also be the reason why she demonstrates a relatively large overlap, as language similarity facilitates acquisition of TEs. This is also evident from the examples found in table 5, demonstrating that her overlap does contain several cognates and form-similar words. A large proportion of her vocabularies, however, are not overlapping and as such her domains are to some extent distributed.

5.3 Are there instances of code-switching in Emma's utterances and if so, what role do the code-switching instances play?

Emma displays instances of code-switching in both English and Norwegian. Tables 6 and 7 demonstrate that she uses more content words in her code-switching than function words. Her code-switching in Norwegian is of a relatively small amount, and she uses it according to two patterns: to change language according to interlocutor or as a lexical gap filling strategy. In the first pattern, as seen in (3), she demonstrates a clear distinction between interlocutors and language use consistent with research demonstrating the role of the interlocutor in relation to language use (Lanza, 1992).

(3)　　　　　**I not know**. (talking to her mum)　　　　　2;8.20

(N: jeg vet ikke)

ja.

(E: yes)

den var ikke igjen der.

(E: it was not left there / it was not there again)

æ skal bare ordne den her.

(E: I am just going to fix this (one))

 ja, jeg har.

(E: yes, I have)

nei.

(E: no)

I like to daddy come in. (talking to her mum)

(N: Jeg vil at pappa skal komme inn)

I like to daddy come in.

(N: Jeg vil at pappa skal komme inn)

Here it has also been found how children are sensitive to the appropriate language to use when and with whom, something that is also evident with Emma who immediately switches from Norwegian only to English only when speaking with her mother. The second pattern she uses can be seen in examples (4)-(6), where she clearly uses code-switching as a lexical gap filling strategy.

(4)   Jeg skal hente dem.   2;8.7

(E: I will get them.)

 Det er <**trim shoes**>.

(E: They are trim shoes)

Jeg skal hente dem.

(E: I will get them)

Du må hjelpe å finne andre.

(E: You must help find the other (one))

Jeg finn bare en joggesko.

(E: I can only find one trim shoe)

(5)   Og en rød ballong her.   2;8.20

(E: And a red balloon here)

Den her andre ballong <**popped**>

(E: The other balloon (here) popped)

Den andre ballong <**popped**>

(E: The othe balloon popped)

(6) Æ vet ikke. 2;9.11

(E: I don't know)

<A reindeer> ?

(N: Et reinsdyr?)

Ka det der er for nokke?

(E: What is that?)

Her English word seems to be more readily at hand before she later uses the appropriate Norwegian word. In other research this has been attributed to the role of language activation and the competition between lexical items, where research demonstrated that both languages become activated in word recognition tasks, and how the resulting word choice might also depend on proficiency and dominance (DeAnda, 2016). It could also be set in connection with research demonstrating how translation equivalents seem to have a facilitative role in lexical access (Poulin-Dubois et al, 2018), where the dominant language might be activated first, and then the right translation equivalent in the other language is found afterwards (DeAnda, 2016). Here one could ask whether using the English, more readily at hand language supported her at the momentarily loss of the Norwegian word, and where the facilitative effect of translation equivalents led her to the Norwegian word *joggesko* used later. Emma's code-switching according to interlocutor and use of English words like in the examples above seem to support the CP. Here she changes between English and Norwegian monolingual language modes according to interlocutor. Her use of code-switching in these instances seems to have a supportive role as a communication strategy. The fact that there is such a small occurrence of code-switching in her Norwegian could support the implications of the CP, as her domains, home and kindergarten, and her interlocutors in the recordings call for a monolingual language mode. This would entail that the activation of the language in question should be larger than the language not required by the domain or interlocutor (Grosjean, 2015).

This pattern is also demonstrated in her code-switching in English. Emma demonstrates a lot of code-switching in her English, in fact more code-switching in her English than she does in her Norwegian. This could also potentially fit the pattern of accessibility in relation to TEs and dominance discussed above, where the Norwegian word is more readily at hand and therefore used first. That she uses Norwegian words to support communication, as in the examples (7) and (8), in the same way as she uses English words to support communication in

Norwegian leads to the question of the occurrence of code-switching could be a demonstration of language balance.

<table>
<tr><td>(7)</td><td>No. Go out!</td><td>2;8.5</td></tr>
<tr><td></td><td>(N: Nei. Gå ut!)</td><td></td></tr>
<tr><td></td><td>Du (?) need to find.</td><td></td></tr>
<tr><td></td><td>(N: Du må finne) I can hide you.</td><td></td></tr>
<tr><td></td><td>(N: Jeg kan gjemme deg)</td><td></td></tr>
<tr><td></td><td>……</td><td></td></tr>
</table>

<table>
<tr><td>(8)</td><td>Those trucks gonna drive.</td><td>2;8.5</td></tr>
<tr><td></td><td>(N: De lastebilene skal kjøre)</td><td></td></tr>
<tr><td></td><td>Du (?) need to drive this truck, (du) (?) Daddy.</td><td></td></tr>
<tr><td></td><td>(N: Du må kjøre denne lastebilen, du pappa)</td><td></td></tr>
<tr><td></td><td>You need to drive that.</td><td></td></tr>
<tr><td></td><td>(N: Du må kjøre den)</td><td></td></tr>
<tr><td></td><td>I gonna drive ...</td><td></td></tr>
<tr><td></td><td>(N: Jeg skal kjøre…)</td><td></td></tr>
<tr><td></td><td>Du (?) need to drive den her (?) truck.</td><td></td></tr>
<tr><td></td><td>(N: Du må kjøre denne lastebilen)</td><td></td></tr>
<tr><td></td><td>Yellow</td><td></td></tr>
<tr><td></td><td>(N: Gul)</td><td></td></tr>
<tr><td></td><td>and this big truck is so big!</td><td></td></tr>
<tr><td></td><td>(N: Og denne lastebilen er så stor!)</td><td></td></tr>
<tr><td></td><td>this truck is so big.</td><td></td></tr>
<tr><td></td><td>(N: Denne lastebilen er så stor)</td><td></td></tr>
</table>

As both her languages cannot be dominant, could the bidirectional pattern of her code-switching, of which TE that seems to be more readily at hand, demonstrate that her languages are balanced? This would, however, not fit the discrepancy between amount of code-switching used in her English compared to amount of code-switching that occurs in her Norwegian. Such asymmetries in amount have usually been attributed to language dominance, where, typically,

such asymmetry follows a code-switching of the dominant language into the weaker language (Nicoladis & Genesee, 1997). This asymmetry is surprising considering her MLUs and TTRs suggest her to be a balanced bilingual, where in fact her English expressive skills seem to be the more lexically diverse.

The contexts Emma finds herself in in either language, is in a monolingual language mode. Something that is not only demonstrated in her shift from Norwegian to English when speaking with her mother but is also demonstrated in examples (9) and (10), where she telephones with her friends.

(9)   MOT: Emma, can you call your friend Emma?   2;8.17

(N: Emma, kan du ringe vennen din Emma?)

EM: Yes.

(N: Ja)

EM: Heia Emma, er du ...

(E: Hey Emma, are you…)

MOT: no, she only speaks English.

(N: Nei, hun snakker bare Engelsk)

EM: huh?

MOT: remember Emma only speaks English.

(N: Husk at Emma bare snakker Engelsk)

MOT: you have to speak English.

(N: Du må snakke Engelsk)

EM: Heia, Emma, is du

(E: Hey Emma, is you)

EM: is du (?) home, Emma?

(N: er du hjemme Emma?)

(10)   That ringed.   2;8.17

(N: Det ringte)

Heia (.) heia.

(E: Hey)

Maybe that is [/] maybe that is Angus?

(N: Kanskje det er/ kanskje det er Angus?)

Heia, Angus!

(E: Hey Angus!)

Er du home, Angus?

(E: Are you home, Angus?)

Ja bye bye, Angus?

(E: Yes bye bye Angus?)

Ja.

(E: Yes)

Now have I ringed Angus.

(N: Nå har jeg ringt Angus)

Heia (.) heia Adrian!

(E: Hey. Hey Adrian!)

Is dokker hjemme, ja.

(E: Are you at home, yes.)

Hadet.

(E: Bye)

When she is talking to her friends, she code-switches from English to Norwegian. Her mother then asks her to call Emma, and when Emma starts speaking Norwegian her mother reminds her that Emma only talks English. Hereafter Emma shifts back to English when talking in the phone. This is a good example of how Emma encouraged to and is used to switch between English only to Norwegian only and vice versa, which again supports how her language mode is predominantly monolingual. However, example (10) demonstrates that when she later talks with her friend Angus, she code-switches in her Norwegian. This could suggest that her English monolingual mode has been "disturbed" by her Norwegian conversations, so she needs some time to get back into her monolingual language mode. It also demonstrates that although her language at home is English, she naturally switches to Norwegian when she is speaking with her friends. This is something she seemingly does automatically and would fit the implications of the CP, where she most likely speaks Norwegian to her friends in kindergarten. As such she separates her language use pattern after interlocutor. What is interesting is how this habit follows into her games at home even when her interlocutor is not present. This could potentially reflect a cognitive aspect of the domain specific pattern of language use, as this is not something that is required from the context, she finds herself in at the time. The increase in code-switching

when making phone calls with her friends could demonstrate a confusion of contexts, she is in an English monolingual language mode, however, with her friends she seemingly seems to be in a Norwegian monolingual language mode, and her combination of those domains could be why her code-switching increases. Indeed, when using the language mode continuum of the CP, this could be explained as she is moving from a monolingual language mode to a more centered language mode along the continuum, where both languages are equally active. The transition from the monolingual language mode over to a heightened activation of her other language could perhaps explain this increase in code-switching. This has also been pointed out by Grosjean (1997) that even when in a monolingual language mode, the other language is never completely deactivated.

Emma's contexts call for monolingual language modes then, and where she finds herself along the language mode continuum potentially influences amount of code-switching. Her predominantly monolingual mode is something the CP predicts will lead to minimal code-switching. Why then the amount of code-switching in her English of such relatively large size is puzzling. As mentioned, research finding such asymmetrical patterns usually also ascribe such asymmetrical pattern of code-switching to language dominance, where the dominant language has more influence on the weaker language, than the weaker language has on the dominant language in terms of code-switching (Paradis & Nicoladis, 2007). However, these results are rather puzzling as her MLU's suggest her English and Norwegian to be balanced, or English to be the more expressively varied and dominant language, if ever so slightly. According to this result of code-switching, however, Norwegian seems to be her more dominant language. This could, as previously mentioned, indicate that she is in a transition phase, like the CP emphasize: dominance across domains can change. Perhaps increased exposure to Norwegian in her Kindergarten has started to change dominance roles for the two languages. Although, it also emphasizes that different domains can have different dominant languages, depending on the need of a given language in different domains. In Emma's instances the role of the interlocutor would here be the modulating factor, as she is at home in a domain where she usually uses her English language. That her Norwegian then is used to such degrees is interesting and puzzling. Another interesting aspect of her code-switching as a communicative supportive function, is her lexical gap filling. As has been demonstrated with other research on code-switching, the interesting with her code-switching is not that she lacks those words, as she has TE's for most of the words she uses in instances of code-switching. This entails that she

does know the majority of words she code-switches. This could perhaps be explained through recent research into language activation, where dominance might influence which language and word is most readily at hand (DeAnda, 2016). This could suggest that language proficiency and dominance will outweigh the cues given from the situation, given her monolingual language mode. Importantly, it also demonstrates how she competently uses both her languages to communicate effectively around the age of 2. An additional note that is of relevance here, is the focus on her code-switching of this thesis. Because the investigation is looking at pragmatic use of code-switching in relation to vocabulary and the CP, any morphosyntactic aspects of code-switching has not been investigated. This will have excluded other kinds of code-switching utterances.

The fact that Emma's code-switching is asymmetrical, is possibly not quite in accordance with the predictions from the CP. Here her code-switching is expected to be at a minimal due to the confined contexts the languages are used in and with restrictions regarding interlocutors. The language mode would suggest minimal amount of code-switching, as the language in the given situation should be the most active one. The CP does allow for both languages to be active, however, given a monolingual language mode, the "dominant" language should be the language required from the situation. However, here it must be pointed out that using the CP as theoretical background has implications for how code-switching is defined and treated. The CP implies a sociolinguistic paradigm for how to look at and interpret code-switching. Here all instances of words or utterances from one language over into the other language is being treated as instances of pragmatical code-switching. Whereas other paradigms functionally separate between code/language mixing and code-switching. There code/language mixing is a more unconscious, structurally founded process rather than the conscious, pragmatic process of mixing words and utterances as a communication strategy (see Stell & Yakpo, 2015).

5.4 Do the characteristics of Emma's vocabularies reflect the differences in Emma's language exposure and language use patterns? Are her vocabularies distributed (home vs. daycare)?

The characteristics of Emma's expressive language skills do seem then to reflect her language exposure and language use pattern. Her MLUs and TTRs indicate balance, something

her language background could also account for, as she seemingly has similar amount of exposure to both languages over time at this point of data collection. Even if this exposure is bulked at different points in time. Her language use pattern is monolingual language use for different situations and interlocutors, this is also evident from her code-switching. She code-switches mainly as a change of interlocutor or to fill lexical gaps. However, despite the bidirectionality of her code-switching, it is asymmetrical. She code-switches more in her English than in her Norwegian.

As she demonstrates a relatively large overlap between her languages, especially for action words, her vocabularies are not as distributed as would be predicted from the CP. However, whether kindergarten and home can be defined as separate or shared domains can be questioned. In addition, the Norwegian-English language pair similarity is also an influential factor on the degree of overlap, which further makes it difficult to draw reliable conclusions as to which degree her vocabularies are distributed.

5.5 How do Emma's expressive lexical skills compare to those of monolingual peers in each language?

5.5.1 MLUS and TTRs

Emma's MLUs and TTRs, see tables 8 -10, do not seem to be that different from that of monolingual peers. She has much the same length, as many words, in her utterances as them and further seems to be as advanced in her utterances, lexically diverse, in her utterances as them. An interesting point here could be to what extent any results could have been different if Emma were compared to girls. Several studies have emphasized how girls and boys sometimes demonstrate differing language development, where boys sometimes demonstrate a slower language development (De Houwer, 2009). There are other questions, however, regarding this way of comparing her TTRs. That is to what extent the average of other averages is a reliable way look into how advanced their languages are. Here other tests are probably more reliable in establishing a basic insight into someone's language skills. In addition, these comparisons are not statistical analyses. As previously mentioned, recordings are reliable in many ways, they do not, however, necessarily reflect the full potential of the candidate. As we can see from the overview 1A (Appendix 1A), there are large variations in number of words produced in total. That is another possible confounding variable here, that the higher scores come from recordings

with low, or lower, number of total words used, which could also influence how high the TTR would be, as there would not be that many words to build an average from. Additionally, as the CP emphasizes, words are often connected to domains, not only necessarily home vs daycare, but it could also be connected to situations (Grosjean, 2015). Helping in the garden would require a different vocabulary than playing with dolls or stuffed animals, thus the situations of Emma's recordings give us a good glimpse of her language skills, but most likely do not reflect her total language skills or knowledge. Which would also apply to the monolinguals. From these glimpses though, she seems to produce as long utterances and as lexically diverse utterances as her monolingual peers.

5.5.2 Vocabulary scores

In Figure 3 I showed how Emma's separate vocabularies are smaller than those of Shem and Ole. This is in accordance with previous research, where bilingual vocabularies are shown to be smaller when the languages are compared separately. However, when compared across languages, in total vocabulary and, potentially, conceptual vocabulary, the bilingual vocabulary(-ies) are not necessarily less than those of monolinguals (Unsworth, 2013). This is also demonstrated here, where Emma's total vocabulary of 597 words is not only comparable to monolingual peers, but also actually larger than Ole's, of 522, and Shem's, of 464. Additionally, Emma's conceptual vocabulary is comparable to Shem's vocabulary. These comparisons highlight what has been emphasized by many researchers of bilingual vocabulary acquisition (De Houwer, 2009, Grosjean, 2015): due to the distributed characteristics of a bilingual's vocabulary, to get a clear image of the language level and language skills of the bilingual, both languages need to be considered. Furthermore, as emphasized by Bosch & Ramon-Casas (2014) language pair studied also needs to be considered when looking into and comparing TSVs and TCVs, as different language pairs can influence level of conceptual vocabulary.

Thus, Emma follows the same developmental trajectory as her monolingual peers, with a comparable vocabulary size once both her vocabularies are considered. As her monolingual peers, she demonstrates a noun prevalence in her vocabularies and demonstrates the same noun prevalence in distribution of lexical items used. There are some qualitative differences, where Emma's difference between noun and verb use in Norwegian is small compared to that of Ole. This is questioned to be related to her overlap, since the languages are similar, and she has

demonstrated to have almost as large an overlap for verbs as she does for nouns. These results are in accordance with the CP, which predicts her vocabulary to comparable to a monolingual level once both her languages are considered.

6 Conclusion

Emma demonstrates a relative balance between her languages, attributed to her exposure pattern of one year at home with minority language exposure and one year in kindergarten with majority language exposure. This suggests relatively similar amount of exposure to each language. Her balance between the languages is something that can be accommodated by the CP as her background of exposure is also relatively balanced. However, here the CP can be speculated to predict that eventually her majority language, Norwegian, might develop to become the dominant language, as she will use this language in more domains and with more people than her minority language, English.

Two of her language characteristics though are difficult to interpret according to the CP. Her amount of overlap is relatively large, where she has a larger overlap for action words rather than content words. This could be accommodated by the CP, which emphasizes that some basic overlap is to be expected. In that respect action words might be more universal than object words. However, a large overlap would indicate that her domains are shared, something that is difficult to establish. Her domains are home and kindergarten. Although these might be considered separate domains due to location, language and people involved, activities might still be the same as at home. This could then also be considered a shared domain. In addition, there is a confounding variable that might also influence amount of overlap. That of language similarities. A larger overlap is to be expected from languages that are similar, such as Norwegian and English. So, in this respect it is difficult to draw any assumptions as to whether her overlap could be explained through separate or shared domains in relation to the CP or whether her overlap is due to language similarities.

The other characteristic is that of code-switching. Here the CP would predict small amounts of code-switching has Emma finds herself in monolingual language modes for both languages. Emma does find herself in monolingual language modes. The CP further predicts that any instances of code-switching should be to support communication. Emma's instances of code-switching are largely to support communication. However, her asymmetric code-

switching between languages indicates other factors that influence amount of code-switching other than what the CP would predict.

Emma's vocabulary (-ies) compared with monolingual peers further demonstrates what the CP, and other research, would predict. When her languages are compared separately, her vocabularies are smaller than those of monolingual peers. However, when her total vocabulary is compared, her vocabulary is not only comparable, but also larger than that of her monolingual peers. This is in accordance with the predictions of the CP as it emphasizes that the distributed nature of vocabularies will lead to discrepancies when vocabularies are compared to monolinguals separately. Whereas when the total vocabulary knowledge is compared, it is not smaller.

The case study of Emma's language characteristics then demonstrates several of the implications of the CP regarding vocabulary acquisition in young, preschool children. It is clear that the CP makes an important theoretical background regarding young bilingual vocabulary acquisition as well as for adult bilingual vocabulary acquisition. However, results from this case study also suggests that there are other factors that are crucial in understating this complex process as well. Bilingual vocabulary acquisition is a complex process and researching it involves considering many aspects of both cognitive and contextual aspects. When investigating bilingual vocabulary acquisition, one must: "…take into account the many underlying phenomena that characterize a person's bilingualism. The CP should be one of them." (Grosjean, 2015, p. 84).

Works cited

Anderssen, M. & Bentzen, K. (2018). Different Outcomes in the Acquisition of Residual V2 and Do-Support in Three Norwegian-English Bilinguals: Cross-Linguistic Influence, Dominance and Structural Ambiguity. *Frontiers in Psychology*, 9, 2130–2130. https://doi.org/10.3389/fpsyg.2018.02130

Bail, A., Morini, G., & Newman, R.S. (2015). Look at the gato! Code-switching in speech to toddlers. *Journal of Child Language*, 42(5), 1073–1101. https://doi.org/10.1017/S0305000914000695

Bentzen, K. (2000). *I like not it like du like it!: A case study of language transfer in bilingual first language acquisition*/[Master thesis, University of Tromsø]

Bialystok, E., Luk, G., Peets, K.F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism* (Cambridge, England), 13(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bornstein, M.H., Cote, L.R., Maital, S., Painter, K., Park, S-Y., Pascual, L., Pecheux, M-G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-Linguistic Analysis of Vocabulary in Young Children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, 75(4), 1115–1139. https://doi.org/10.1111/j.1467-8624.2004.00729.x

Bornstein, M.H., Putnick, D.L., & De Houwer, A. (2006). Child vocabulary across the second year: Stability and continuity for reporter comparisons and a cumulative score. *First Language*, 26(3), 299–316. https://doi.org/10.1177/0142723706059238

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary. *International Journal of Behavioral Development*, 38(4), 317–322. https://doi.org/10.1177/0165025414532559

Brandeker, M. & Thordardottir, E. (2015). Language exposure in bilingual toddlers: performance on nonword repetition and lexical tasks. *American Journal of Speech-Language Pathology*, 24(2), 126–138. https://doi.org/10.1044/2015_AJSLP-13-0106

Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism* (Cambridge, England), 16(1), 32–48. https://doi.org/10.1017/S1366728912000120

Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences - PNAS*, 114(34), 9032–9037. https://doi.org/10.1073/pnas.1703220114

Byers-Heinlein, K. and Werker, J.F. (2009). Monolingual, bilingual, trilingual: infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, 12(5), 815–823. https://doi.org/10.1111/j.1467-7687.2009.00902.x

Byers-Heinlein, K. and Werker, J.F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*, 128(3), 407–416. https://doi.org/10.1016/j.cognition.2013.05.010

Chiaro, D. (2009) Cultural divide or unifying factor? Humorous talk in the interaction of bilingual, cross-cultural couples. In Norrick, N.R., & Chiaro, D. (eds) (2009). *Humor in Interaction* (Vol. 182). John Benjamins Publishing Company.

Childers, J.B. & Tomasello, M. (2006). Are Nouns Easier to Learn Than Verbs? Three Experimental Studies. In Hirsh-Pasek, K. & Golinkoff, R. (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). New York: Oxford University Press.

Chondrogianni, V. (2018). Child L2 acquisition. In Miller, D., Bayram, F., Rothman, J., & Serratrice, L. (Eds.) (2018). *Bilingual Cognition and La*nguage. *The state of the science across its subfields.* (p. 103 - 125) (Vol. 54). John Benjamins Publishing Company.

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish-English bilinguals from 22 to 30 months: implications for assessment. *Journal of speech, language, and hearing research: JSLHR*, 56(5), 1637–1649. https://doi.org/10.1044/1092-4388(2013/11-0044)

Correia, L., & Flores, C. (2017). The Role of Input Factors in the Lexical Development of European Portuguese as a Heritage Language in Portuguese–German Bilingual Speakers. *Languages* (Basel), 2(4), 30. https://doi.org/10.3390/languages2040030

David, A., & Wei, L. (2008). Individual Differences in the Lexical Development of French–English Bilingual Children. *International Journal of Bilingual Education and Bilingualism*, 11(5), 598. https://doi.org/10.2167/beb478.0

DeAnda, S., Poulin-Dubois, D., Zesiger, P. & Friend, M. (2016). Lexical Processing and Organization in Bilingual First Language Acquisition: Guiding Future Research. *Psychological Bulletin*, 142(6), 655–667. https://doi.org/10.1037/bul0000042

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, 28(3), 411-424. http://doi:10.1017/S0142716407070221

De Houwer, A. (2009). *Bilingual first language acquisition* (Vol. 2, pp. XVII, 412). Multilingual Matters.

De Houwer, A., Bornstein, M.H., & Putnick, L. (2014). A Bilingual-Monolingual Comparison of Young Children's Vocabulary Size: Evidence from Comprehension and Production. *Applied Psycholinguistics*, 35(6), 1189–1211. https://doi.org/10.1017/S0142716412000744

Eilers, R.E., Pearson, B.Z., & Cobo-Lewis, A.B. (2006). Social Factors in Bilingual Development: The Miami Experience. In McCardle, P. & Hoff, E. (Eds.) (2006) *Childhood Bilingualism: Research on Infancy through School Age* (pp. 68–90). Channel View Publications. https://doi.org/10.21832/9781853598715-006

Gentner, D. (1982) Why are nouns learned before verbs: linguistic relativity versus natural partitioning. In *Language Development: Language, Thought and Culture* (Vol. 2) (Kuczaj, S.A., ed.), In pp. 301–334, Lawrence Erlbaum Associates

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544–564). New York: Oxford University Press.

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, 12(10), 397–403. https://doi.org/10.1016/j.tics.2008.07.003

Gopnik, A. And Choi, S. (1995). Names, relational words and cognitive development in English and Korean speakers: nouns are not always learned before verbs. In Tomasello, M., & Merriman, W. E. (1995). *Beyond names for things: young children's acquisition of verbs* (pp. VI, 421). L. Erlbaum.

Grosjean, F. (1997). The Bilingual Individual. (1997). *Interpreting*, 2, 163 – 187.

Grosjean, F. (2015). The Complementarity Principle and its impact on processing, acquisition, and dominance. In Silva-Corvalán, C., & Treffers-Daller, J. (2015). *Language*

Dominance in Bilinguals. Cambridge University Press.
https://doi.org/10.1017/CBO9781107375345

Harris, M. (2004). First words. In Oates, J., & Grayson, A. (Eds.). (2004). *Cognitive and language development in children.* Blackwell Publishing; Open University Press.

Hoff, E., & Core, C. (2015). What Clinicians Need to Know about Bilingual Development. *Seminars in Speech and Language,* 17(2), 089–099. https://doi.org/10.1055/s-0035-1549104

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hoff, E., Rumiche, R., Burridge, A., Ribot, K.M. & Welsh, S.N. (2014). Expressive vocabulary development in children from bilingual and monolingual homes: A longitudinal study from two to four years. *Early Childhood Research Quarterly*, 29(4), 433-444. https://doi.org/10.1016/j.ecresq.2014.04.012.

Johnson, C.E., & Lancaster, P. (1998). The Development of More Than One Phonology: A Case Study of a Norwegian-English Bilingual Child. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, 2(3), 265–300. https://doi.org/10.1177/136700699800200302

Junker, D.A. & Stockman, I.J. (2002). Expressive Vocabulary of German-English Bilingual Toddlers. *American Journal of Speech-Language Pathology*, 11(4), 381–394. https://doi.org/10.1044/1058-0360(2002/042)

Kaushanskaya, M., & Crespo, K. (2019). Does Exposure to Code-Switching Influence Language Performance in Bilingual Children? *Child Development*, 90(3), 708–718. https://doi.org/10.1111/cdev.13235

Lanza, E. (1992). Can bilingual two-year-olds code-switch*? Journal of Child Language*, 19(3), 633–658. https://doi.org/10.1017/S0305000900011600

Legacy, J., Reider, J., Crivello, C., Kuzyk, O., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2017). Dog or chien? Translation equivalents in the receptive and expressive vocabularies of young French–English bilinguals. *Journal of Child Language*, 44(4), 881–904. https://doi.org/10.1017/S0305000916000295

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: method, theory and practice.* Cambridge University Press.

Meisel, J. (2006). The bilingual child. In Bhatia, T. K., & Ritchie, W. C. (eds) (2006). *The handbook of bilingualism.* Blackwell Pub.

Montanari, E.G., Abel, R., Graßer, B., & Tschudinovski, L. (2018). Do bilinguals create two different sets of vocabulary for two domains? *Linguistic Approaches to Bilingualism*, 8(4), 502–522. https://doi.org/10.1075/lab.16021.mon

Nicoladis, E. (1998). First clues to the existence of two input languages: Pragmatic and lexical differentiation in a bilingual child. *Bilingualism* (Cambridge, England), 1(2), 105–116. https://doi.org/10.1017/S1366728998000236

Nicoladis, E., & Genesee, F. (1997). Language Development in Preschool Bilingual Children. *Journal of Speech-Language Pathology and Audiology*, 21(4), 258–270.

Nicoladis, E. & Secco, G. (2000). The role of a child's productive vocabulary in the language choice of a bilingual family. *First Language*, 20(58), 003–028. https://doi.org/10.1177/014272370002005801

Oller, D. K., Eilers, R. E., Urbano, R., & Cobo-Lewis, A. B. (1997). Development of precursors to speech in infants exposed to two languages. Journal of Child Language, 27, 407–425. https://doi.org/10.1017/S0305000997003097

Otwinowska, A., & Szewczyk, J.M. (2019). The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *International Journal of Bilingual Education and Bilingualism*, 22(8), 974–991. https://doi.org/10.1080/13670050.2017.1325834

Paradis, J. & Nicoladis, E. (2007). The Influence of Dominance and Sociolinguistic Context on Bilingual Preschoolers' Language Choice. *International Journal of Bilingual Education and Bilingualism*, 10(3), 277–297. https://doi.org/10.2167/beb444.0

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43, 93-120. doi:10.1111/j.1467-1770.1993.tb00174.x

Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J.. (2013). Lexical access and vocabulary development in very young bilinguals. *The International Journal of*

Bilingualism: *Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, 17(1), 57–70. https://doi.org/10.1177/1367006911431198

Poulin-Dubois, D., Kuzyk, O., Legacy, J., Zesiger, P., & Friend, M. (2018). Translation equivalents facilitate lexical access in very young bilinguals. *Bilingualism* (Cambridge, England), 21(4), 856–866. https://doi.org/10.1017/S1366728917000657

Qiu, C., & Winsler, A. (2017). Language use in a 'one parent-one language' Mandarin-English bilingual family: noun versus verb use and language mixing compared to maternal perception. *International Journal of Bilingual Education and Bilingualism*, 20(3), 272–291. https://doi.org/10.1080/13670050.2015.1044935

Quick, A.E., Lieven, E., Tomasello, M., & Backus, A. (2018). Constructively combining languages: The use of code-mixing in German-English bilingual child language acquisition. *Linguistic Approaches to Bilingualism*, 8(3), 393–409. https://doi.org/10.1075/lab.17008.qui

Ribot, K.M., & Hoff, E. (2014). "¿Cómo estas?" "I'm good." Conversational code-switching is related to profiles of expressive and receptive proficiency in Spanish-English bilingual toddlers. *International Journal of Behavioral Development*, 38(4), 333–341. https://doi.org/10.1177/0165025414533225

Saffran, J., Werker, J. A., & Werner, L. A. (2006). 'The infant's auditory world: Hearing, speech, and the beginnings of language. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, Volume 2: Cognition, perception, and language* (pp. 59-108). Hoboken, N. J.: John Wiley & Sons.

Schelletter, C. (2002). The effect of form similarity on bilingual children's lexical development. *Bilingualism* (Cambridge, England), 5(2), 93–107. https://doi.org/10.1017/S1366728902000214

Smithson, L., Paradis, J., & Nicoladis, E. (2014). Bilingualism and receptive vocabulary achievement: Could sociocultural context make a difference? *Bilingualism* (Cambridge, England), 17(4), 810–821. https://doi.org/10.1017/S1366728913000813

Smolak, E., de Anda, S., Enriquez, B., Poulin-Dubois, D., & Friend, M. (2020). Code-switching in young bilingual toddlers: A longitudinal, cross-language investigation. *Bilingualism* (Cambridge, England), 23(3), 500–518. https://doi.org/10.1017/S1366728919000257

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology, 32*(3), 492-504. https://doi:10.1037/0012-1649.32.3.492

Thordardottir, E. (2011). The relationship between bilingual exposure and vocabulary development. *The International Journal of Bilingualism : Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, 15(4), 426–445. https://doi.org/10.1177/1367006911403202

Thordardottir, E., Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, 4, 1-21. https://doi:101080/14769670500215647

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *The International Journal of Bilingualism:* Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior, 23(1), 236–255. https://doi.org/10.1177/1367006917722418

Unsworth, S. (2013). Current Issues in Multilingual First language Acquisition. *Annual Review of Applied Linguistics 33*, 21 – 50.

Vagh, S. B., Pan, B.A., & Mancilla-Martinez, J. (2009). Measuring Growth in Bilingual and Monolingual Childrens English Productive Vocabulary Development: The Utility of Combining Parent and Teacher Report. *Child Development*, 80(5), 1545–1563. https://doi.org/10.1111/j.1467-8624.2009.01350.x

Von Holzen, K. & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, 113(4), 569–586. https://doi.org/10.1016/j.jecp.2012.08.001

Walla, D. (2017). *Code mixing in early bilingual acquisition: dominance, language modes, and discourse strategies. A case study of bilingual acquisition of Norwegian and English*. [Master thesis, University of Tromsø]

Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H.. (2013). Are Nouns Learned Before Verbs? Infants Provide Insight Into a Long-Standing Debate. Child Development Perspectives, 7(3), 155–159. https://doi.org/10.1111/cdep.12032

Yamamoto, M. (2001). Language use in interlingual families: a Japanese-English sociolinguistic study (Vol. 30). *Multilingual Matters*.

Yip, V., & Matthews, S. (2006). Assessing Language Dominance in Bilingual Acquisition: A Case for Mean Length Utterance Differentials. *Language Assessment Quarterly*, 3(2), 97–116. https://doi.org/10.1207/s15434311laq0302_2

Yow, W.Q., Tan, J.S.H, & Flynn, S. (2018). Code-switching as a marker of linguistic competence in bilingual children. *Bilingualism* (Cambridge, England), 21(5), 1075–1090. https://doi.org/10.1017/S1366728917000335

Appendix 1A and 1B

Appendix 1A Overview language characteristics Emma's English

	TNW	NDW	TTR	TNN	NDN	TNV	NDV
2;7.14	660	186	0,282	86	43	178	34
2;8.5	1647	257	0,156	208	65	360	45
2,8.17	1364	280	0,205	292	78	288	51
2;9.2	1275	245	0,192	144	53	292	47
2;9.23	798	207	0,259	64	51	186	41
2;10.8	267	99	0,371	34	14	50	16

TNW= total number of words used, NDW= number of different words used, TTR=type/token ratio, TNN=total number of nouns used, NDN= number of different nouns used, TNV=total number of verbs used, NDV=number of different verbs used.

Appendix 1B Overview language characteristics Emma's Norwegian

	TNW	NDW	TTR	TNN	NDN	TNV	NDV
2;7.21	869	169	0,194	111	38	223	37
2;8.7	642	149	0,232	67	31	165	30
2,8.20	1332	219	0,164	113	43	287	38
2;9.11	1132	208	0,184	79	48	249	46
2;9.25	1331	226	0,17	99	52	296	46
2;10.9	2099	279	0,133	156	72	472	48

TNW= total number of words used, NDW= number of different words used, TTR=type/token ratio, TNN=total number of nouns used, NDN= number of different nouns used, TNV=total number of verbs used, NDV=number of different verbs used.

Appendix 2A and 2B

Appendix 2A Verb overlap

Total overlap verbs							
Eng	Norw	Eng	Norw	Eng	Norw	Cognates	
ask	be	say	si	wake	våkne	like	like
be	være	show	vise	push	trykke	see	se
can	kan	take	ta	tell	si	fall	falle
catch	fange	wait	vente	tie	knyte	pack	pakke
do	gjøre	want	vil	shop	handle	find	finne
drink	drikke	wash	vaske	think	tro	help	hjelpe
drive	kjøre	talk	prate	turn	snu	set	sette
eat	spise	build	bygge	fancy	like	sit	sitte
sleep	sove	braid	flette	fit	passe	will	vil
fix	ordne	come	komme	try	prøve	dance	danse
get	få	jump	hoppe	read	lese	fly	fly
have	ha	close	lukke	need	trenge	ring	ringe
wear	ha på	hide	gjemme	play	leke	match	matche
know	vite	buy	kjøpe	watch	kikke	hold	holde
make	lage	call	ringe	look	lete	hang	henge
move	flytte	lie	ligge			put	putte

Appendix 2B Noun overlap

Total overlap nouns									
Eng	Norw	Eng	Norw	Eng	Norw	Cognates			
year	år	ceiling	tak	horse	hest	hand	hånd	dinosaur	dinosaur
bear	bamse	comb	børste	kitty	kattepus	music	musikk	egg	egg
car	bil	cow	ku	pig	gris	thing	ting	hat	hatt
crown	krone	doll	dukke	toe	tå	baby	baby	pizza	pizza
daddy	pappa	door	dør	button	knapp	ball	ball	sock	sokk
dog	hund	dress	kjole	coat	kåpe	cake	kake	Box	boks
eye	øye	hair	hår	jacket	jakke	finger	finger		
grandma	bestemor	hole	hull	floor	gulv	fish	fisk		
grandpa	bestefar	house	hus	sweater	genser	foot	fot		
								Trim	
mommy	mamma	lid	lokk	bird	fugl	hammer	hammer	shoes	Joggesko
puzzle	puslespill	nose	nese	undies	truse	man	mann		
rabbit	kanin	brush	børste	boy	gutt	room	rom		
road	vei	bandage	plaster	mitten	vott	teletubby	teletubby		
train	tog	money	penger	noise	Lyd	balloon	ballong		
block	brikke	chicken	kylling	Pooh Bear	Ole Brumm	book	bok		
calf	kalv	flower	blomst	Piglet	Nasse Nøff	cat	katt		

Appendix 3A and 3B

Appendix 3A Overview language characteristics Shem

	TNW	NDW	TTR	TNN	NDN	TNV	NDV
2;7.10	1802	352	0,195	247	105	341	61
2;8.3	2199	387	0,176	236	108	465	70
2;8.20	1152	257	0,223	137	74	255	51
2;9.1	935	277	0,296	119	63	202	44
2;9.27	1106	273	0,247	137	72	237	45
2;10.2	933	236	0,282	116	65	203	55

TNW= total number of words used, NDW= number of different words used, TTR=type/token ratio, TNN=total number of nouns used, NDN= number of different nouns used, TNV=total number of verbs used, NDV=number of different verbs used.

Appendix 3B Overview language characteristics Ole

	TNW	NDW	TTR	TNN	NDN	TNV	NDV
2;7.20	1761	350	0,199	233	91	417	70
2;8.5	2255	368	0,163	316	102	476	61
2;8.24	2499	425	0,17	333	111	583	83
2;9.15	2100	390	0,186	249	106	456	76
2;10.0	1977	328	0,166	217	82	504	59
2;11.23	1082	266	0,246	151	79	249	53

TNW= total number of words used, NDW= number of different words used, TTR=type/token ratio, TNN=total number of nouns used, NDN= number of different nouns used, TNV=total number of verbs used, NDV=number of different verbs used.