UiT
THE ARCTIC UNIVERSITY OF NORWAY

Faculty of Humanities, Social Sciences and Education

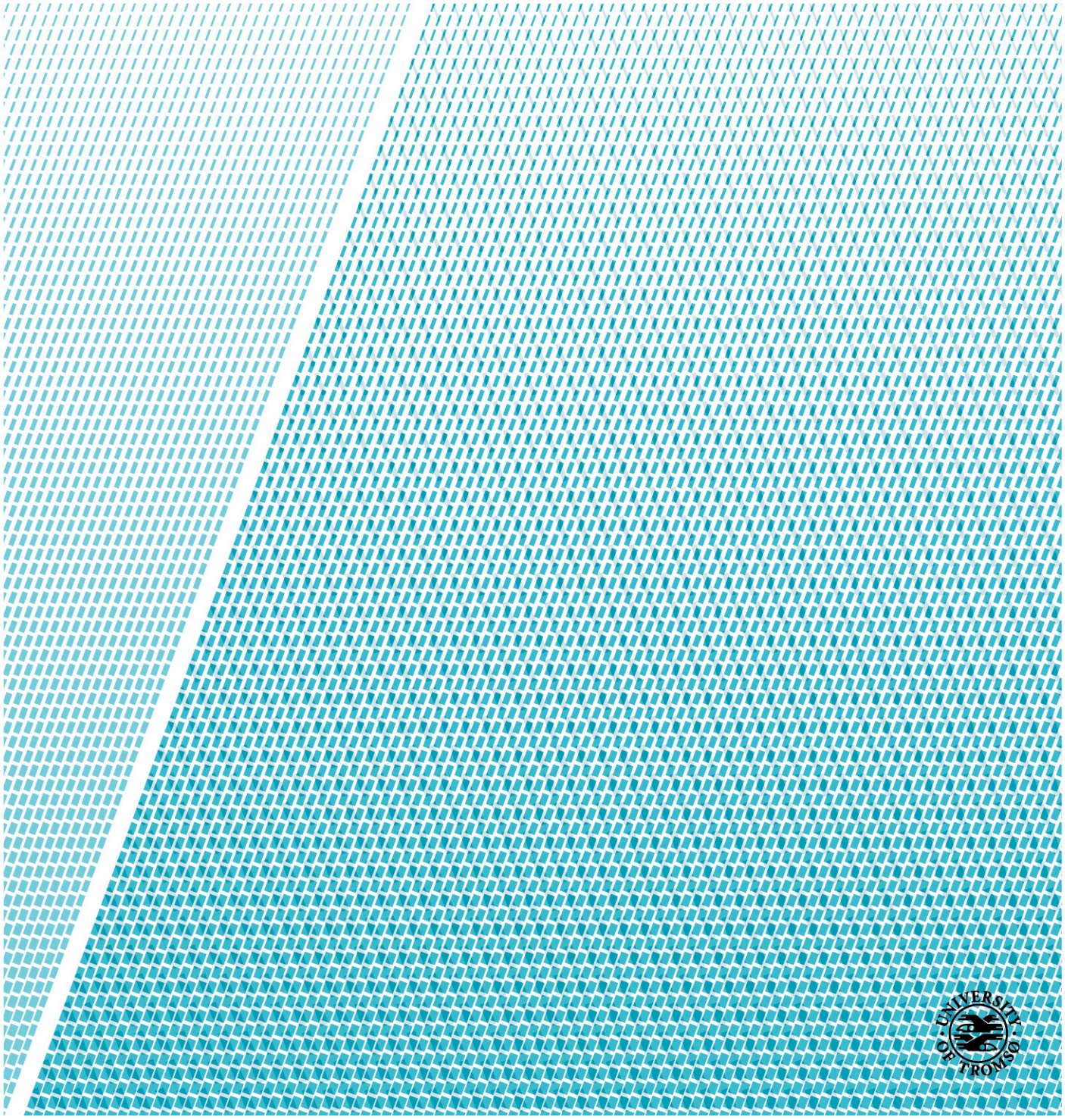
Code mixing in early bilingual acquisition: dominance, language modes, and discourse strategies

A case study of bilingual acquisition of Norwegian and English

—

Dianna Walla

LIN-3990 Master's thesis in Theoretical Linguistics - May 2017



Abstract

One of the main questions that arises in cases where children code switch or code mix often in their early production is whether mixing can be considered a sign of confusion, of competency, or of neither. Earlier research on language mixing in bilingual first language acquisition often pointed to the idea of a "unitary language system," with code mixing being considered a sign of confusion. More recent and systematic research contradicts the claim of the hypothesis, however. More recently, researchers have proposed that bilingual children use code mixing as a way to fill lexical gaps where translation equivalents are not yet present.

This study examines the code mixing of a bilingual girl, Hedda, growing up in northern Norway and acquiring Norwegian and English from birth. Hedda's father is American and consistently speaks English with her, while Hedda's Norwegian mother consistently addresses her in Norwegian. The parents speak Norwegian with each other. Thus, although her father is a consistent and ample source of English input, Norwegian is clearly the dominant language in Hedda's surroundings. The thesis uses corpus data collected between the ages of 2;3-3;3 to examine Hedda's mixing habits when conversing with native speakers of both Norwegian and English in order to compare her code mixing habits in each language. The data show that Hedda code mixes extensively when communicating with English speakers, while code mixing is almost absent in Norwegian contexts. Hedda's mastery of English lags behind her Norwegian competence (the latter is comparable to that of her monolingual Norwegian peers). Norwegian is clearly her dominant language. Moreover, her day-to-day life rarely brings her into contact with monolingual English contexts, making a monolingual English mode a rare occurrence for her. Based on this, the thesis examines possible contributing factors to account for Hedda's code mixing patterns.

Acknowledgments

I am deeply indebted to the people who helped me make this thesis a reality. Thank you to my supervisor Kristine Bentzen, who became not just a supervisor but also a mentor to me. Her feedback was invaluable throughout the process and it was always delivered in an encouraging way, and I can confidently say that this thesis is better for it. I want to thank the whole LAVA research group for their support and camaraderie, but especially Marit Westergaard, our research group leader, and Merete Anderssen, leader of the work package that the Hedda corpus belongs to, and the one who got me involved in the corpus project in the first place – it is not every day you get handed such an incredible opportunity. Taking part in the regular LAVA seminars and events has enriched my experience as a graduate student at UiT immeasurably and I am so grateful to have been able to participate.

Massive thanks goes to Hedda, the subject of this study, and her parents, without whom this thesis would never have been possible. They welcomed me into their home from the very beginning and I am eternally grateful for their participation in the project, and for the extra joy they have brought into my life over the past year and a half as I have gotten to know Hedda. Thank you.

And lastly, thanks go to my incredibly supportive friends and family. Thank you to my parents for being the proudest parents anyone could ever hope for, to the friends who have always been encouraging of my academic goals, and especially to my unbelievably wonderful husband Chris, who was willing to follow me all the way to the Arctic so I could pursue those goals. He is one-of-a-kind.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 A case study of bilingual acquisition of Norwegian and English

This thesis uses a case study to explore specific issues in bilingual first language acquisition. Based on recordings of a child acquiring Norwegian and English from birth, issues involving code switching and code mixing as well as potential cross-linguistic influence are examined systematically. The recordings were collected in a corpus and some additional information aided the research, such as parental observations. The child in question, Hedda, is very dominant in Norwegian, producing very few utterances containing solely English for the majority of the recording period covered here. This thesis aims to examine Hedda's mixing habits when conversing with native English speakers in order to determine if any patterns emerge. Based on that, and using research on mixing in other cases of BFLA, I will seek to examine possible contributing factors to account for Hedda's strong language dominance in order to contribute further to the study of child language acquisition and bilingualism.

Because code switching and code mixing are not random in adult speech, but rather are systematic, this study seeks to approach mixing in child language with that in mind. While language mixing in child language does not necessarily resemble adult code switching and code mixing practices, it is also not entirely random. There are several things to take into account when examining Hedda's mixing. What shape does mixing take? Single words from one language borrowed into an utterance in the other language, or switching languages mid-utterance? Is there a correspondence with phrase structure in the syntax? Does mixing occur in one direction more than the other (i.e. from Norwegian into English, or from English into Norwegian)? Is there a difference in frequency of mixing between content/lexical words or functional grammar words? And lastly, is who Hedda is conversing with a factor in her mixing behavior? These are some of the questions that are considered for the analysis of the data. Language dominance, input quantities, and discourse strategies will prove highly relevant.

1.2 Thesis structure

An overview of previous research in bilingual first language acquisition is presented in

Chapter 2. Some notably early studies are described, followed by more detailed descriptions of research that pertains to specific areas of bilingual first language acquisition that I will explore through Hedda's data. A brief outline of some of the relevant differences between Norwegian and English is also provided.

Chapter 3 covers the methodology of this study, which is part of a larger ongoing research project. Methods for data collection, transcription, and analysis are outlined, and a discussion of some of the potential methodological issues is included.

The preliminary data obtained from fourteen English recording sessions and four Norwegian recording sessions is presented in Chapter 4, with a discussion of Hedda's different language production patterns in each of these contexts. Hedda's language dominance is assessed through a number of measures and her code mixing habits are discussed with her dominance in mind. The influence of discourse strategies is also considered, and evidence for cross-linguistic influence in her language production is presented. Discussion of these phenomena is carried out throughout the chapter.

Chapter 5 concludes this thesis, summarizing the findings and placing them within the bigger picture of bilingual first language acquisition research as a whole.

2 The state of the art

This chapter will provide an overview of some of the research and theory on bilingual first language acquisition (BFLA) relevant to this thesis. While BFLA has not been explored as thoroughly as monolingual first language acquisition, today there exists a relatively rich variety of research on the topic which forms the foundation of the work being carried out today. The first section of this chapter will provide a brief overview of some of the landmark studies in BFLA that continue to be referenced in the literature, and subsequently research covering more specific aspects of BFLA which are relevant for this thesis will be covered.

2.1 Bilingual language acquisition: a brief overview

The field of bilingual acquisition research has only just begun to come into its own in recent decades, growing exponentially in the past few decades, but several pioneering studies in the field of child bilingualism have provided unique insight into the early direction of bilingual acquisition research, even if, as De Houwer (1990) notes, the methodologies are not always ideal by modern standards.

Ronjat (1913) is a study of his son Louis's acquisition of French and German, and it was a lengthy and detailed piece of work for its time. Ronjat was a big proponent of the one person, one language (1P/1L) approach to child-directed speech. De Houwer (1990, 2009) provides a nice overview of Ronjat's work.

Werner Leopold was a German-American phonetician who chronicled the acquisition of German and English by his daughter Hildegard. She was being raised in the United States with one parent speaking German to her (Leopold) and one parent speaking English to her (his wife, Hildegard's mother). Leopold's documentation and analysis of Hildegard's bilingual first language acquisition spans four volumes, published between 1939 and 1949. This work had a major impact on child language acquisition research being conducted several decades later, including the rather well-known paper by Volterra and Taeschner from 1978, "The acquisition and development of language by bilingual children." De Houwer (1990) is also able to point to potential methodological flaws in the case of Leopold's study, but it is an important piece of work in the BFLA literature nonetheless.

Burling (1959) wrote about his son's acquisition of English and Garo. While this was technically not a case of bilingual first language acquisition (as defined below), Burling's son's exposure to Garo began at the age of one year, four months, which makes it a particularly early case of early second language acquisition. Due to the unique situation of Burling's son – born in the United States and exposed to English for the first sixteen months of his life before the family relocated to India, when exposure to Garo joined his English input – it was a most interesting study for its time. Burling notes that "for the greater part of the time we were there, [Stephen's] Garo was significantly more fluent than his English."

According to Annick De Houwer (2009), Swain (1976) was the first to introduce the term *bilingual first language acquisition* (henceforth BFLA), which was later picked up by Meisel (1989). De Houwer (1990, 2003, 2007, 2009) has followed this example, using this term to refer to children who have been exposed to two languages from birth ("from birth" being the key criterion). BFLA is sometimes abbreviated in the literature as 2L1 acquisition, and it can be distinguished from other language acquisition contexts such as monolingual first language acquisition (henceforth MFLA) as well as early second language acquisition (henceforth ESLA). De Houwer (1990, 2009) defines the crucial criterion for BFLA as having been regularly exposed to two or more languages from birth: "BFLA children are learning two first languages. There is no chronological difference between the two languages in terms of when the children started to hear them" (De Houwer 2009, 2). There has not always been a general consensus about where the line between BFLA and ESLA stands, and in much of the earlier work done on early bilingualism, researchers have written about cases they considered to be BFLA, which may be defined as ESLA but not as BFLA by De Houwer's definition. I will follow De Houwer's definition of BFLA in this thesis.

Genesee (2016) presents a nice historical overview of some of the existing bilingual research. He emphasizes that much of the work that has historically been done on bilingual speakers has been based on the assumption that monolingualism is the norm, and bilingual speakers have been compared to them within that mindset (with the idea that an ideal bilingual is two monolinguals in one being prevalent). Today, there is fortunately wide acknowledgement that this is not the case in the global sense - in fact, being bi- or multilingual is actually *more* common than monolingualism, and it would be folly to compare bilinguals to a monolingual standard of acquisition without taking into account the unique properties of acquiring two or more languages at once or other environmental or situational factors. De Houwer (2009) has

also written on this topic: “Somehow, people expect bilinguals to be two monolinguals in one – and not just any monolinguals, but monolinguals with highly advanced language skills.” Grosjean (2008), too, has emphasized what he calls “the wholistic view of bilingualism,” which “proposes that the bilingual is an integrated whole which cannot be easily decomposed into two separate parts. The bilingual is *not* the sum of two complete or incomplete monolinguals; rather, he or she has a unique and specific linguistic configuration.”

Where there is variation found in cases of monolingual first language acquisition, there is much greater variation found in cases where more than one language is being acquired, due to much wider variation in the input, differing social attitudes towards language and bilingualism, and other factors. On this theme, Genesee (2016) adds “that bilingualism is not categorically better, harder, or more advantageous than monolingualism; and that not all bilinguals are the same.”

In other words, oversimplification of the realities of language acquisition is not a useful tool, whether monolingual or bilingual acquisition is the focus. We cannot say that monolingual or bilingual acquisition is unilaterally good or bad, or that one is inherently better than the other. Rather a consideration of the multifaceted nature of language acquisition will serve us better, whether we are researchers, parents, or educators.

Grosjean (1982) outlines a number of commonly occurring instances in bilingual acquisition, listed in roughly chronological order of appearance:

a) An initial mixed language stage
b) The separation of the two systems and increasing awareness of bilingualism
c) Influence of one language on the other when the linguistic environment favors one of the languages
d) Avoidance of difficult words and structures in the weaker language
e) Final separation of the sound and grammatical systems but enduring influence of the dominant language on the nondominant language in some domains

While I take some issue with (a), which seems to fall in line with the stages proposed by Volterra and Taeschner (1978) (see Section 2.4), I will agree that (b)-(e) are all well-attested in the existing literature. These tendencies are also present in Hedda’s case, as we will see later in Chapter 4.

I will turn now to the more specific issues and research within the field of BFLA most relevant to this thesis. The effect of input, the concept of language modes, the practices of code switching and code mixing, and a short overview of relevant differences between Norwegian and English will all be covered.

2.2 Input effects

We begin with input, for it is being exposed to input in two different languages that is one of the most important factors leading to bilingual acquisition. Input has been an often-studied topic in language acquisition research in recent years – whether children are acquiring more than one language or just one, it has become clear that input plays a significant role in language acquisition, both in terms of the *quantity* and the *quality* of input (Unsworth 2014). In monolingual children, it has been shown that the amount of input has a direct effect on children's language development using measures such as vocabulary size or mean length of utterance (MLU) (Hoff 2003; Huttenlocher et al. 1991). When children are acquiring two or more languages from birth or from a very young age, the question of input becomes even more complicated, as the same number of hours in a day are divided by two languages. Several new questions emerge when examining the role of input in bilingual contexts: how much input does the child receive in each of their languages? In which contexts do they receive input in each language and how do those contexts differ? What effect does this have on the type of vocabulary they develop in each language? And more recently, researchers have begun to examine more closely the difference between relative input (i.e. measures of input quantity based on the proportion of input in one language to the other) and absolute input (i.e. absolute measures of input in each language, by word count or some other means). It has become clear that this distinction makes a critical difference when comparing language input to child output, for example (Grüter et al. 2014).

In many BFLA contexts (as opposed to ESLA contexts), a child's parents speak two different native languages (De Houwer 2009, Slavkov 2015). The one person, one language (1P/1L) approach has been common; in this method, each parent speaks (usually) their native language to the child in an attempt to provide balanced input in each language. Both Slavkov and De Houwer (2003, 2007) have noted, however, that the 1P/1L model doesn't necessarily always lead to the desired outcome of a child who is an active bilingual productively using

both languages. There are a number of reasons for this, but De Houwer has shown that when one parent's native language is the majority language of society and the other parent's language is a minority language, this model is most successful if the parents speak the minority language with each other – that is, when not speaking to the child (see Section 2.2.3 below for an overview of De Houwer's large-scale study that led to this conclusion). This non-child-directed speech still makes a difference because it still makes up a proportion of the child's input and probably also has an impact on establishing language modes in the home. To provide an example, if a child growing up in Norway has one parent who speaks English to the child and one parent who speaks Norwegian to the child, the 1P/1L model is more likely to be successful if the parents also speak English, the minority language, with each other, rather than Norwegian. As we will see with Hedda, this is not the case in her situation.

2.2.1 Relative versus absolute input

Grüter et al. (2014) explores this question regarding relative input measures versus absolute input measures. Their study examined the different outcomes when comparing these two measures with regard to bilingual children's vocabulary development and online processing. In their words, "[i]t is not straightforward to determine how to operationalize the construct of 'input', and thereby, to capture meaningful variability in the features of those environments." In other words: how exactly does one measure input? One typical approach has been to use *relative* measures of input, by estimating proportions or percentages of input in each language for children in bilingual environments. For a bilingual child, one might propose that approximately 75% of input is in Language A while the other 25% of input is in Language B. (A child growing up monolingual in Language A could be said to receive 100% of their input in Language A.) This type of relative measure is unable to account for differences in *absolute* input, however. Knowing that a child receives 75% of their input in Language A tells us nothing about the average number of words a child hears in a day. Grüter et al. (2014) developed a set of hypothetical data in order to illustrate this point:

Table 1 *Summary of input and output measures in hypothetical sample (Grüter et al. 2014)*

	Julian	Luis	Alexa	Kamila
% time with Spanish speakers	50	67	50	50
% time with English speakers	50	33	50	50
Total Spanish words heard per day	5,000	10,000	10,000	10,000
Total English words heard per day	5,000	5,000	5,000	10,000
Total Spanish words in productive vocabulary	100	200	200	200
Total English words in productive vocabulary	100	100	100	200

In Table 1, rows 1-2 indicate the relative amount of input in each language, rows 3-4 indicate the absolute amount of input in each language, and rows 5-6 indicate output in terms of productive vocabulary. The table makes it very easy to see that comparing relative input with absolute measures of child output obscures the relationship between input and output, while comparing absolute input with child output yields a direct correlation. Given that this is hypothetical data, even if we ignore for the time being the rows concerning productive vocabulary (rows 5-6), this example still makes it very obvious that relative input measures are unable to capture the differences present in the absolute measures of input.

One of the reasons that relative input in the form of parental reporting has so commonly been the measure used in studies of input in child language acquisition is that absolute input is logistically a challenge to measure. New technology is helping researchers to overcome this challenge, however. Both Grüter et al. (2014) and De Houwer (2014) make note of the LENA™ (Language Environment Analysis) system as a tool for determining absolute input. The LENA system makes use of a small, lightweight recorder, worn in a pocket on the child's clothing. The recorder is capable of recording for up to 16 hours, so it is able to capture whole days at a time. The software is then able to provide estimates for adult word counts. While this system is unable to provide information about the quality of input, the estimates are incredibly useful for measuring the quantity of absolute input. The software is unable to distinguish between languages at this point in time, so in order for it to be useful for measuring the absolute input in each language in cases of bilingual acquisition, having a system for determining which language is being spoken to the child (such as a 1P/1L situation) may be necessary, and the software's results should be carefully double checked in order to ensure accuracy.

In order to lend support to their proposal regarding relative versus absolute input frequency, Grüter et al. provide preliminary data from two ongoing studies which make use of the LENA system in order to acquire direct estimates of input (see Grüter et al. 2014 for a full explanation of how they accounted for the bilingual environment in order to obtain direct estimates of input for both languages involved in the study).

The first set of analyses involves using LENA to acquire direct estimates of relative input (as opposed to indirect measures, like parental surveys in which the parents try to provide an estimate based on memory and perception alone). In other words, the study functions as a check on the accuracy of parental reporting as a measure of relative input. While the data is still preliminary, the early results are illuminating in that the analyses "so far show that estimates of relative English exposure based on parental interview reveal a different picture than more direct measures of relative amount of English heard by the child based on LENA recordings of caregiver-child interaction over a typical day at home." Even though this data is still dealing with input in terms of proportion (i.e. relative input), the proportion itself is obtained by making use of the recordings themselves, rather than a parental estimate. In other words, parental estimates of relative input may be even less reliable than we thought.

The second study actually derives estimates of absolute input in terms of the "absolute word count" (henceforth AWC). The preliminary results focus on two children, both of whom "had about 60% of their daily interactions in Spanish, according to parent report." The measures of absolute input for each child revealed that for one, the input density is much greater than for the other: an average of nearly 1,500 words per hour compared to 500 words per hour. This underlines the point made by Grüter et al. (2014) as far as the inability of relative input measures to account for absolute input.

While these studies are still ongoing and the results are only preliminary, what they reveal about the nature of relative and absolute estimates of input is likely to make a great difference in research on input in the future.

2.2.2 Changes in input

While there is obviously variation in the input between children, it can be difficult to control for environmental factors when seeking to conduct studies comparing variation in the input. For this reason, situations where variation in the input is present within a single child's timeline can provide unique insight. A particularly illuminating study of input effects was carried out by Slavkov (2015), regarding the effects of an international trip (and thus a period of an entirely different linguistic environment) on child language production. The child, Sophie, was growing up in Canada and had been exposed to English and Bulgarian from birth. Through the use of parental diary notes and recordings of spontaneous speech, Slavkov was able to document the child's production in the periods of time preceding a 10-day trip to Bulgaria, during the trip, and returning home to Canada. Sophie's mother was a native speaker of English and her father a native speaker of Bulgarian, and they employed the 1P/1L strategy when addressing her, each using their native language. Until Sophie started half-day daycare outside the home (which was an English environment), she seemed relatively balanced as a developing bilingual. Once daycare was introduced, however, Sophie's production in English increased and production in Bulgarian decreased.

The trip to Bulgaria turned all this on its head. Sophie and her father spent ten days in Bulgaria when Sophie's age was 2;3, but her native English speaking mother couldn't come on the trip. This meant Sophie's English input temporarily disappeared. The day before the trip, and for each day of the trip, Sophie's father recorded spontaneous interactions between them for periods of 7-20 minutes (11 recordings total). The goal of the recordings was to be able to compare Sophie's language production in Bulgarian just before the trip to the period of the trip itself. The difference in the proportion of each language in Sophie's language production is immense. The day before the trip, the proportion of Sophie's utterances that were in Bulgarian rather than English was 13%. By Day 7, the proportion of Bulgarian was near 100%.

This study also points to the likely relevance of parental discourse strategies. Slavkov demonstrates how Sophie's father consciously used discourse strategies that would establish a monolingual (Bulgarian) context. This worked very well to elicit Bulgarian utterances from Sophie until she started daycare, and the amount of input she received in each language suddenly skewed much more strongly toward English. At this point, "[t]he diary indicates that

in this period direct requests for translation and other strategies aimed as maintaining a monolingual discourse context were not successful as the child began to employ her own strategies of insisting on a bilingual context." As a result, the father began using strategies that were less aggressive in establishing a monolingual discourse context. While Slavkov makes no claims about whether parental discourse strategies are effective in a child becoming a productive bilingual, the diary notes and the data seem to suggest that their efficacy is reliant to a degree on the amount of input – and perhaps there is even an input threshold, below which parental discourse strategies attempting to establish a monolingual context won't be as effective. This, however, is mere speculation for now.

2.2.3 Input patterns in the home

Finally, an impressively large-scale study of input patterns was carried out by De Houwer (2003, 2007). She collected data in the Flanders region of Belgium, where Dutch is the sole official language. Data collection was carried out via questionnaires that were sent home with schoolchildren. The self-reporting questionnaire format undoubtedly presents certain disadvantages, but the tradeoff is the ability to conduct a study on a much larger scale than more detailed investigations generally allow for. The study drew data from 1,942 questionnaires from bilingual families, and for the purposes of the study she defined a bilingual family as one in which "at least one family member spoke a single language other than Dutch at home." De Houwer used the results of the questionnaires to examine which languages were used by different family members within the home.

Given the scale of the study, she could compare different input contexts within the home: where "Language X" refers to any language other than Dutch, she was able to compare five different environment types, where: both parents spoke only Language X, both parents spoke Language X and one parent also spoke Dutch, both parents spoke both Language X and Dutch, both parents spoke Dutch and one parent also spoke Language X, or both parents spoke only Dutch. These different environments are outlined in the table below:

Table 2 *Based on De Houwer (2007)*

Parent 1	Parent 2
X	X
X & Dutch	X
X & Dutch	X & Dutch
Dutch	X & Dutch
Dutch	Dutch

In the table there are two instances of environments where both parents speak one language only (Language X and Language X; Dutch and Dutch) and there are two instances of environments where one parent is monolingual and one parent is bilingual (X + Dutch and X; Dutch and X + Dutch); nonetheless these are worth distinguishing because Dutch is the majority language, and the children receive Dutch input outside the home as well (overall more input in Dutch than in Language X).

Most importantly, the analysis of the data showed that there was a great deal of variation in how successfully children became productive speakers of the minority language, Language X. Depending on the input environment, children in two-parent homes had a success rate between 35.70% and 96.92%, a massive difference. De Houwer found that the input environment did make a significant difference: the most successful input patterns were homes where both parents spoke only Language X at home (with a success rate of 96.92%) and homes where both parents spoke Language X and only one parent also spoke Dutch (93.42%), meaning the parents would speak Language X (the minority language) together. 1P/1L model homes where one parent spoke Language X and the other spoke Dutch at home had a success rate of 74.24%, indicating that the conventional wisdom that elevates the effectiveness of this approach is flawed. The least successful input environment was homes where both parents spoke Dutch and only one parent spoke Language X, with a success rate of 35.70%.

Even given the limitations of working with data collected in a survey format, the clear implication here is that the amount of input in each language makes a very great difference,

and in a monolingual social environment where children are learning the majority language (spoken by the monolingual society) and a minority language, it is the minority language that is at risk of not being acquired due to too little input.

2.3 Language modes and the complementarity principle

Another important factor to consider when conducting research on bilingualism is that of language modes. François Grosjean has written extensively on the importance of language modes when conducting research on bi- and multilingualism, particularly with regard to language acquisition. For bilinguals he conceptualizes a continuum with a fully monolingual mode at one end and a fully bilingual mode at the other, where "[a] mode is a state of activation of the bilingual's languages and language processing mechanisms" (Grosjean 1998). In other words, depending on the context and who the bilingual is speaking to, each language will be at a corresponding level of activation or deactivation. Interacting with a monolingual speaker of a language (let's call it Language A) will mean Language A is activated for the bilingual. Language B, the bilingual's other language which the interlocutor does not speak, will be less activated. These levels of activation can vary based on many different variables, and there is considerable variation among bilingual speakers as well, particularly when it comes to levels of dominance. Grosjean (2008) notes that in cases where bilinguals are highly dominant in one of their languages, deactivating the dominant language can be difficult even in situations where the speaker starts out in a monolingual mode with an interlocutor who only speaks the weaker language. In his words, "the weaker language will simply not be developed enough to allow them to stay in a monolingual mode" (Grosjean 2008).

In Hedda's case (as we will see in later chapters), she is highly dominant in Norwegian, which certainly affects her ability (or inability) to maintain a monolingual mode in her weaker language, English. The fact that she rarely encounters situations that could be described as being in a monolingual English mode will also be of crucial importance.

In a survey of existing BFLA research at the time, De Houwer (1990) notes while assessing the methodologies of several studies whether or not the studies had accounted for language mode. When discussing Leopold (1970), she mentions more than once that Leopold's data collection methods could have had an effect on the language choice of the child in question,

his daughter Hildegard. “Leopold was the only data collector. He was a fluent bilingual, and spoke both English and German in the home…Whenever data were gathered, Leopold was always present, with or without Hildegard’s mother being there as well.” In other words, although Leopold usually spoke German to Hildegard, he spoke English with her mother or when there were other English speakers around, so Hildegard was aware of his bilingualism. De Houwer makes the point that this could affect Hildegard’s language choice. Or, to put it in Grosjean’s terms, the presence of Leopold as data collector would theoretically be sufficient to establish a bilingual mode for Hildegard, even if the goal was to record an English interaction between Hildegard and her mother. We will return to this point later for Hedda’s case, as it is particularly relevant.

Grosjean’s work has also focused on what he calls the complementarity principle, which goes hand-in-hand with the concept of language modes: “[b]ilinguals usually acquire and use their languages for different purposes, in different domains of life, with different people. Different aspects of life often require different languages” (Grosjean 2008). The idea underlying this principle is that bilinguals typically do not need to use *both* of their languages in *every* realm of life. There may be some environments or situations in which both languages will be engaged, but often there are realms (home, school, or work, perhaps) where only one of the languages is necessary and engaged. The complementarity principle has implications for concepts such as “fluency” – bilingual speakers may appear to be fluent in one of their languages, say Language A, in one context, but in a different context they’ll lack the fluency and vocabulary of monolingual speakers because they typically use Language B in that context. The point Grosjean wants to stress through the complementarity principle is that bilinguals should be considered holistically as speakers in their own right, not as two monolinguals in one person. This point has been picked up and stressed by other researchers since he first made it, including Annick De Houwer (2009). It is also important to note that the complementarity principle does *not* suggest that different realms of life must make use of only one language or the other – there is a third option, which is that some realms may make use of both languages, creating a bilingual context that would establish a bilingual mode (such as the home of a family where the parents speak two different languages to their child, for example).

2.4 Code switching and code mixing

One commonly observed phenomenon in bilingual acquisition is code switching or code mixing during the first few years of language production. While the terminology has not always been used entirely consistently, *code switching* tends to refer to inter-sentential switching between languages within a single discourse, while *code mixing* often refers to mixing two or more languages within one sentence or utterance (Yow et al. 2016). Other researchers have distinguished these two terms in different ways – or sometimes not at all – but I will use them as just described within this thesis: code switching will refer to inter-sentential language switches within a single discourse, while code mixing will refer to intra-sentential mixing (within a single utterance).

One of the main questions that arises in cases where children code switch or code mix often in their early production (particularly the latter) is whether switching and mixing can be considered a sign of confusion, of competency, or of neither. Earlier research on language mixing in bilingual first language acquisition often pointed to the idea of a unitary language system (Swain 1972, Volterra & Taeschner 1978), with researchers arguing that "language mixing by infant bilinguals is interpreted as a sign of the child's lack of awareness of actually dealing with two languages" (Lanza 1992). In other words, the unitary language system hypothesis proposes that bilingual children begin developing only one linguistic system, unaware that their two languages are two separate systems. Volterra & Taeschner (1978) proposed a three-stage model of bilingual acquisition in which the first stage seemed to support the unitary language system hypothesis. More recent and systematic research contradicts the claim of the hypothesis, however (see Genesee 1989 and De Houwer 2009 for an overview). According to Yow et al. (2016), there is evidence that bilingual children under the age of 3;6 use code mixing as a strategy to fill lexical gaps (somewhat like a bilingual overextension of lexical items). Genesee (2008) and Green et al. (2012) have reported similar findings. Other research suggests that when children are older, the use of code switching and mixing is a sign of having high competence in both languages (De Houwer 2009), being systematic and conforming to the grammars of one or both languages. There is also evidence that rates of code switching in child language show a correlation with the rates of code switching in parental input, though as we will see in later chapters, this is not the case for Hedda.

2.4.1 Mixing in previous BFLA case studies

Lanza's (1992) systematic study of code mixing by a child acquiring English and Norwegian from birth is especially relevant to this case study. There are many similarities between the situation of Siri, the child Lanza worked with, and Hedda's situation: one Norwegian parent and one American parent, both bilingual[*], who use a 1P/1L approach when addressing their child. Siri was also growing up in Norway, although in a different region than Hedda. The biggest difference is that Siri's parents spoke English together (the minority language), unlike Hedda's parents who use Norwegian when speaking to each other (the majority language), which very likely makes a significant difference with regard to the proportions of input in each language that each child received.

A few of the key points when it comes to Lanza's study include the finding that when Siri's rates of grammatical mixing were compared with her rate of lexical mixing (i.e. functors versus content words) for each parent, differences emerged. "Siri engages in more lexical mixing with her Norwegian-speaking father than with her English-speaking mother" (Lanza 1992). Additionally, her grammatical mixing was asymmetrical in that while she mixed Norwegian functional morphemes into her English when speaking with her (English-speaking) mother, she did not mix English functional morphemes into her Norwegian when speaking with her (Norwegian-speaking) father, only lexical ones. Lanza points to this as evidence of Siri's dominance in Norwegian. When it comes to Siri's lexical mixing, however, she actually did *more* lexical mixing when speaking to her father in Norwegian, her dominant language, than when speaking with her mother in English. In order to explain this, Lanza points to context and the discourse strategies employed by each parent. Siri's mother used these strategies to evoke a more monolingual context, while her father, whether intentionally or not, created a more bilingual context in the discourse strategies he employed.

There are three particularly relevant strategies: the **minimal grasp strategy**, the **expressed guess strategy**, and the **repetition strategy**. The minimal grasp strategy involves using

[*] A note on my use of the term 'bilingual': I do not restrict this term to refer only to childhood bilingualism, as I consider it the most straightforward way to refer to any person who has communicative proficiency in two languages, regardless of the age of onset for the L2 (or the second L1, in BFLA contexts). As such, by referring to parents in the literature as bilingual, I do not necessarily mean to imply that they also grew up being exposed to more than one language, only that as adults and as parents, they have two languages at their disposal.

discourse cues like “What?” or “I don’t understand” to attempt to signal that the child has not been understood and should use the other language. Of the strategies named, this is most useful for trying to establish a monolingual, rather than a bilingual, context. The expressed guess strategy is a slightly less effective way of signaling that the parent or interlocutor wishes for the child to use the other language; it involves recasting what the child just said in Language A in the form of a question in Language B. If Hedda used the word *blåbær* in Norwegian in an otherwise English-speaking context, for example, the interlocutor might respond, “Do you mean a blueberry?” in order to communicate that what was just said may not have been understood and that English would be a better language to use in this context. The repetition strategy is the least effective. While quite similar to the expressed guess strategy at first glance, the repetition is not cast in the form of a question, but rather a statement. So in the case of Hedda using the Norwegian word *blåbær* in an English-speaking context, the interlocutor would respond with something like, “a blueberry, yeah.” This would not require a response from the child, unlike the other strategies.

De Houwer’s (1990) case study of Kate, a child acquiring Dutch and English in a BFLA context also presents a systematic analysis of the child’s code mixing habits. The relevant utterances are referred to as “Mixed utterances” by De Houwer, and she extracted these Mixed utterances into a data set and coded them according to several different factors, some of which include whether the utterance was a response or an initiation within the conversation, who the interlocutor was if it was a response (and which language(s) they spoke), and whether the utterance was mainly Dutch (MMD), mainly English (MME), or “Dutlish” (determined by the number of morphemes present from each language). In general, the child’s language choice was determined by two main factors (for monolingual utterances as well as Mixed utterances): the language the interlocutor used to address the child, and the child’s awareness of that interlocutor’s linguistic abilities (i.e. whether they were monolingual or bilingual speakers). This latter point is particularly relevant for studies seeking to analyze code mixing in a bilingual child’s utterances, and is important in Hedda’s case because of my own bilingual identity, which she likely perceived even though I never spoke Norwegian in front of her.

De Houwer found that when it came to Kate’s mixed utterances, the interlocutor she was communicating with was a significant factor. The MMD utterances were “mainly directed at a Dutch speaker” while the MME utterances were “mainly directed at an English speaker” (p.

96). “Dutlish” utterances were spoken to interlocutors who Kate knew were bilingual speakers of Dutch and English much more frequently than they were spoken to monolingual English speakers. In other words, the amount of mixing even within Kate’s mixed utterances was higher when she was speaking with bilingual interlocutors. De Houwer links these results to the question of language choice, and notes that “Mixed utterances cannot be seen as evidence that the bilingual child does not make use of two separate language systems. Instead, they show her creative manipulation of the tools for fluent speech production: a bilingual lexicon, and two closed linguistic rule systems.”

2.5 Cross-linguistic influence

Somewhat related to code switching and code mixing (although a separate phenomenon) is that of cross-linguistic influence (CLI), also referred to in the literature as interference or transfer. As the name suggests, CLI involves the interaction of two or more of a speaker’s languages, with one influencing the other in some way. In the case of CLI in BFLA, this of course presupposes that the child’s two languages are developing as separate (if not entirely autonomous) systems.

In order to distinguish between code mixing and CLI, I turn to Paradis and Genesee (1996). Following their definition, code mixing “indicates an ‘on-line’ interaction between the two languages in performance” and it is not necessarily a sign of systemic interaction. In other words, it has to do with processing and not with competence. What they call interdependence, on the other hand, is defined “as being the systemic influence of the grammar of one language on the grammar of the other language during acquisition,” where their use of the word “systemic” indicates that influence occurs at the representation or competence level. One potential manifestation of interdependence is transfer, in which a grammatical property from one language is incorporated into the other. The level of syntactic complexity the child has reached in each language has an effect on this, with the language in which the child has reached a more advanced stage influencing the syntax of the less well-developed language.

Some definitions of CLI are more restrictive, in that they state that CLI-related transfer occurs when a syntactic construction in a child’s two languages overlap (Müller 1998): in one language, the structure is unambiguous, but in the other language, an ambiguity exists (two syntactic orders may be possible, for example). Transfer occurs when the child makes use of

the unambiguous structure from one language to deal with the ambiguity in the recipient language.

In order to prepare for my own analysis of Hedda's code mixing and potential cross-linguistic influence, a brief description of some of the structural properties of English and Norwegian follow in the next section, as well as an overview of relevant research on the acquisition of these two languages, whether in a monolingual or a bilingual setting.

2.6 An overview of Norwegian and English

Compared with many other BFLA language combinations, Norwegian and English are fairly similar in typology with a large number of cognates. There are nonetheless a number of differences between the two languages that are potential sources of cross-linguistic influence. A short overview of some of the pertinent differences follows here.

Norwegian is a Germanic language with verb-second (V2) word order. There is a three-way gender system for nouns (masculine, feminine, and neuter), and definite nouns are marked with a suffix. English differs from Norwegian in terms of syntax in a few key ways: unlike the other Germanic languages, English is not a V2 language (although it maintains some remnants of its V2 past), and English has no grammatical gender. Definite nouns are marked with a prenominal determiner rather than a suffix.

2.6.1 Verbs in Norwegian and English

Syntactically, V2 means that Norwegian sentences undergo verb movement, where finite verbs always move out of the VP in main clauses; English differs in that main verbs always remain in V, while finite auxiliaries undergo movement. This difference is made explicit when the placement of adverbs is compared:

(1) English: She always **eats** breakfast.
Norwegian: Hun **spiser** alltid frokost.
she eats always breakfast

(2) English: Yesterday she **ate** breakfast.
Norwegian: I går **spiste** hun frokost.
yesterday ate she breakfast

In (1), the verb *eats* follows the adverb because it remains in V, whereas the verb *spiser* in Norwegian precedes the adverb because it has raised to C. Likewise in (2) the English verb remains in V, while the Norwegian verb raises to C, inverting the order of the verb and the subject. This difference can also be seen in question form, where the English version requires the insertion of an auxiliary.

(3) English: **Did** she **eat** breakfast?
Norwegian: **Spiste** hun frokost?
ate she breakfast

The question requires inversion, but as the English verb remains in V, *do*-support supplies an auxiliary, whereas the Norwegian main verb can invert without the aid of an auxiliary. In English, the only main verb which can undergo this kind of inversion is the copula *be* (Santelmann et al. 2002). In this case, the structures of the Norwegian and the English overlap.

(4) English: **Are** you hungry?
Norwegian: **Er** du sulten?
are you hungry

The Norwegian system could be considered less ambiguous than the English given the greater consistency – and though it always requires finite verb movement in a main clause, which might otherwise be a sign of complexity, it could be considered the simpler option when compared with the exceptions and the need for auxiliary support in comparable English sentences. And indeed, research has shown that V2 properties are acquired relatively early in monolingual children acquiring Germanic languages like Norwegian and Swedish, as early as 2;0 (Westergaard 2005, Platzack 1996, Santelmann 1995).

The question that arises in the case of the simultaneous acquisition of two languages with competing systems, such as Norwegian and English, is whether one will be acquired before the other, and if so, is there transfer? (Bentzen 2000)

2.6.2 Other phenomena

As mentioned, Norwegian and English differ on grammatical gender – Norwegian has a three-gender system where English does not have grammatical gender. Research has shown the complete acquisition of gender to be a relatively late phenomenon (Rodina and Westergaard 2015), however (around age 7), so I will not pay very much attention to it in this case study. It will suffice to say that the three-gender system in Norwegian is made up of masculine, feminine, and neuter genders, with masculine being the default gender for new nouns and feminine being marked. In some dialects of Norwegian, including the Tromsø dialect, there is evidence that the feminine gender is being lost (Rodina and Westergaard 2015).

One other difference mentioned was that of definite marking of nouns. Indefinite nouns are marked with a prenominal determiner in both English and Norwegian. Definite nouns are a different matter. English again uses a prenominal determiner, but Norwegian instead makes use of a suffix. Example (5) shows the parallel structure of indefinite nouns, while example (6) shows this difference.

(5) English: **a** book
Norwegian: **en** bok
a book

(6) English: **the** book
Norwegian: bok**en**
book-the

Anderssen & Bentzen (2013)

This difference will not be hugely important in the analysis here, but it is somewhat relevant to the discussion of Hedda's mixing of Norwegian functors into English in Chapter 4.

2.7 Conclusion

This chapter has gone through a great deal of the relevant background information from the existing BFLA literature. While still a relatively new area of study, research in bilingualism has grown exponentially in the last 30 years, so there is fortunately a wide body of work to draw upon. I have provided an overview of some of the early landmark studies in the field, then more specific issues were discussed, such as issues dealing with input and its effect on

BFLA and Grosjean's concept of language modes and his complementarity principle. Code switching and code mixing as well as cross-linguistic influence were discussed, and these will prove particularly relevant for the analysis in Chapter 4, and some relevant information regarding the syntactic differences between Norwegian and English was provided.

3 Methodology

An overview of the methodology for this study will be provided in this chapter. First, Hedda's relevant background information is presented, after which follows a description of the data collection and transcription methods employed. A discussion of the potential issues and shortcomings of this study is also included, with suggestions for ways to overcome some of those shortcomings in the future.

Before proceeding with the overview of the methodology, a note about examples and formatting is in order. The examples presented here are largely left in the format used in the corpus, which involves the use of certain symbols and codes. While these aid the ease of analysis and provide additional information for a researcher, they can be somewhat cumbersome to read and require some explanation. Every line of speech begins with a three-letter code representing a speaker: *CHI for the child (Hedda), *INV for the investigator (the author of this thesis), *MOT for Hedda's mother, and *FAT for her father. The default base language of utterances for the sessions I recorded was English, so utterances where the base language is Norwegian are preceded by the code [- nor], as these are marked. I have removed the code normally associated with mixed words for ease of reading and instead used boldface to indicate code mixing within an utterance: words appearing in boldface belong to the opposite language compared to the rest of the utterance. Maintaining the transcript format also allows me to share information that isn't strictly linguistic, such as pauses, non-verbal vocalizations, and other commentary that was included in the transcript. Pauses are indicated by a period or periods within parentheses, with the number of periods indicating the relative length of the pause: (.), (..), or (...). Non-verbal vocalizations or nonce forms are preceded by the & symbol, and commentary about an utterance may be presented on the line below after the code %com.

3.1 Hedda's case study

Hedda is growing up in northern Norway with a Norwegian mother (native language Norwegian) and an American father (native language English). Both parents are bilingual in Norwegian and English and they use Norwegian when speaking together. Hedda is acquiring the Tromsø dialect of Norwegian and, to an extent, American English. Her father is her primary source of input in English, and prior to the beginning of our recording sessions, he

was the only regular source of English input (and given that the recording sessions were only 2-4 times per month, I could hardly be considered a regular source of input). For the most part, the parents employ the 1P/1L strategy, with the mother addressing Hedda in Norwegian the father addressing her English. Hedda is obviously aware that her father speaks and understands Norwegian, however, and while he only addresses her in English, he accepts responses from Hedda in Norwegian. Exceptions to the 1P/1L strategy occur most often when there are people around who don't speak Norwegian, and the mother frequently uses English with Hedda in these cases. Hedda attends a typical Norwegian-speaking *barnehage*, or daycare, during the daytime. Based on the combination of attending Norwegian-speaking daycare with the home situation, her proportion of input on a daily basis is much higher for Norwegian than for English.

Hedda's parents are both highly educated with an interest in Hedda's linguistic development, and both hold positive attitudes toward bilingualism. They both use Norwegian and English in daily life outside the home. Aside from the decision to employ a 1P/1L strategy, language planning has not played a huge role in the family. As previously mentioned, Hedda's father accepts responses from Hedda in Norwegian, even though he addresses her in English. In fact it is Hedda's mother who is more likely to explicitly tell Hedda that she should be using English with an English-speaking interlocutor (such as the author), although Hedda's father occasionally does this too. Code switching could be said to be used by the parents inside the home, but this behavior was dependent on interlocutor and context. Hedda's father code switched between Norwegian and English when at home with both Hedda and her mother, as he typically spoke English to Hedda and Norwegian to her mother. And when I was present in the home for recordings, both parents spoke English to me, but continued to speak Norwegian to each other when discussing things in the background if I was interacting with Hedda. Additionally, Hedda's mother frequently switched between English and Norwegian utterances when addressing Hedda in my presence during the recording sessions, though the majority of these switches were inter-sentential. As a rule, Hedda's parents did not code mix, and instances of code mixing by the parents found in the transcripts are either cases of explicitly asking for or providing Hedda with translation equivalents as in Example (7a), or repeating one of Hedda's mixed utterances because it was amusing or invoked some other emotional response, as in Example (7b).

(7a) (Session 11, 3;1.5)

*MOT: what are you looking for in this bag?
*CHI: [- nor] gullskoan mine.
gold+shoes my
'My gold shoes.'
*MOT: well yeah and what's that in English?
*CHI: [- nor] æ vet ikke.
I know not
'I don't know.'
*MOT: well (..) **gull** is gold.
*CHI: gold.

(7b) (Session 9, 2;11.17)

*INV: don't need help.
*INV: Hedda doesn't need help.
*CHI: don't **gi** help.
give
*CHI: don't **gi** help.
*MOT: don't **gi** help!
%com: realizing what Hedda is saying.
*MOT: &=laughs is that what you're saying?

Given this background information, we can say a few things about Hedda's situation. In the case of Grosjean's model of language modes (2008), Hedda often finds herself in a monolingual Norwegian mode (at daycare or around Norwegian family members) and a bilingual Norwegian/English mode (at home with both parents), but a monolingual English mode is a truly rare occurrence in her life since her main source of English input and her main conversation partner in English also speaks Norwegian and accepts Norwegian responses. Sporadic trips to countries where English is the main language of communication and the family was surrounded by English speakers may have induced monolingual English modes for short periods of time, but her typical situation rarely allows for that.

As far as Grosjean's (2008) complementarity principle, Hedda has been exposed to both Norwegian and English words relating to home and travel, but daycare is an environment where she hears Norwegian words only. A few of the recording sessions took place shortly after trips to English-speaking countries: there was a 10-day family trip to the United States (that both parents went on) between Sessions 4 and 5 with Session 5 taking place approximately three weeks after the family returned, and Session 10 took place 3 days after returning from a 10-day trip to South Africa (both parents were also on that trip). Additionally, Hedda's American grandmother and her partner, neither of whom speak

Norwegian, visited the family in Norway for one week between Sessions 13 and 14. While the family remained at home in Norway, this nonetheless constituted a sizable shift in Hedda's input situation for the duration of the visit, not only because of the visiting family members who only spoke English, but also because Hedda's mother spoke much more English, including to both Hedda and her father in contexts where she normally wouldn't (and at the expense of Norwegian).

Hedda displays a marked dominance in Norwegian. For a majority of the recordings, few of her utterances contain only English words, even when interacting with English-speaking interlocutors. Many of her utterances in these contexts contain a great deal of code mixing, and a large number of utterances contain only Norwegian, even when speaking with an interlocutor whom she has never heard speak Norwegian. This thesis will seek to examine Hedda's language choice and her mixed utterances in English-speaking contexts with an eye toward her input situation, the establishment of different language modes, the discourse strategies of her interlocutors, and other relevant factors, with the goal of determining which of these factors has an effect and whether there is any interaction between the factors. The following sections will provide an overview of the data collection, the transcription, and tools used for analysis in this case study.

3.2 Data collection

The majority of the data on which this thesis is based comes from naturalistic recordings of Hedda in a home setting. The Language Acquisition, Variation, and Attrition (LAVA) research group at the University of Tromsø – The Arctic University of Norway are carrying out an ongoing project to build a dense corpus of a Norwegian-English bilingual child (Hedda) that can be used for a variety of research projects. This is a new corpus and data collection is ongoing.

The recordings were separated into two main types: Norwegian recordings and English recordings. Due to Hedda's heavy Norwegian dominance, however, the "English" recordings contain a great deal of Norwegian. Nonetheless, I will refer to them here as the Norwegian recordings and the English recordings in order to distinguish between them easily, because the circumstances surrounding them do differ. The Norwegian recordings began when Hedda was at the age of 1;1.30 and were conducted by her mother. A smartphone was used for the

recordings and the data is very naturalistic: there was not necessarily any effort to actively elicit speech from Hedda. Her mother would be present at home for the duration of the recording but might not be actively interacting with Hedda continuously throughout the recording. Her father may also be present during the Norwegian recordings, when he would address Hedda in English as usual. There was also no specified time during the week for these recordings to take place, and they are as a rule much shorter in length than the English recordings. For the purposes of this thesis, I disregard the Norwegian recordings that were conducted prior to the age of 2;3, when the English recordings with an external investigator began, but some of the recordings that took place during the same period of time as the English recordings are relevant.

In contrast, the English recordings began when Hedda was 2;3.10 and were conducted by the author of this thesis. The decision to bring in an external interlocutor/observer for the English recordings was made due to Hedda's knowledge of her father's bilingual identity, and her tendency to use Norwegian when speaking to him. Recordings took place on a specified day and time each week at the child's home and there was a conscious effort to elicit English speech from Hedda. Recording began shortly after my arrival and lasted for the duration of the visit, usually one to two hours. Rather than repeating one or two similar activities on any kind of regular schedule, the recordings feature the participants in free play activities of all kinds, which may include playing games, reading books, playing pretend with Hedda's toy kitchen, or building with Legos. The underlying assumption was that letting Hedda guide the play would encourage her participation and communication. It should be noted that the type of activity or game could influence her language choice, however; some games were given to her by English-speaking family members and Hedda thus learned to play with them in English. Likewise, she tended to discuss books or movies in the language those media used: she was more likely to use English words when talking about the Disney films *Snow White* or *Frozen*, which she watched in English, while she very rarely used English words to talk about the characters or events in the Norwegian story *Dyrene i Hakkebakkeskogen.*

With a few exceptions, the English recordings were conducted on a small portable recorder, unlike the Norwegian recordings. Due to travel and other conflicts, there were occasionally gaps in the recording schedule, so recordings did not always take place every week. It should be noted that with respect to language mode, recording in the child's home means it's impossible to consider the language mode of the environment during the English recordings

as being a truly monolingual English mode – one or both parents were always present at home even if they weren't in the room for periods of the recording, and since both languages were used at home by both parents, this environment naturally evokes a bilingual mode. Hedda in fact had very few opportunities to operate in a monolingual English mode, since she was living in Norway and her main source of English, her father, accepted responses from Hedda in Norwegian (even though he only addressed her in English). Additionally, while I was careful to never speak Norwegian in front of Hedda, I understand Norwegian and could usually understand Hedda's Norwegian utterances. This played an indirect role on the discourse strategies employed. I sometimes used the repetition strategy in responding to her as opposed to one that would indicate a lack of understanding or elicit an English utterance instead, as in the following example:

(8) (Session 5, 2;7.14)
*INV: what is that?
*CHI: xxx.
*CHI: [- nor] pølse!
hot-dog
'Hot dog.'
*INV: is it a hot dog?

In this instance, Hedda answered the initial question with the Norwegian word, and I employed the repetition strategy to keep the conversation flowing. This type of exchange happened often, and I was not always conscious of the fact that this could indicate to Hedda that I understood the Norwegian word, which was not my intent.

The English recordings were transcribed orthographically by the author. The first two English recordings included in this study were transcribed using the Computerized Language Analysis software (CLAN), and the remainder were transcribed using EUDICO Linguistic Annotator software (ELAN) and were later converted to CLAN-compatible files for analysis. In both cases transcription was carried out according to the guidelines for Codes for the Human Analysis of Transcripts (CHAT) format. Due to time constraints, these transcripts were proofread by a native Norwegian speaker, but they have not yet been transcribed by anyone else. Analysis was carried out using CLAN. At the time of writing, only a selection of the large number of recordings had been transcribed; these are used in the analysis. The recordings that are not yet transcribed are nonetheless mentioned because the frequency of recordings meant Hedda usually saw the author several times a month and was very familiar

with her. Out of 36 recordings conducted by the author in total at the time of writing, 14 have been included in the analysis as the core data set; an overview of these is given in Table 3.

Table 3 *An overview of the English recordings used for the analysis.*

Session number	Age	Recording length
1	2;3.10	1:13:32
2	2;4.0	1:37:23
3	2;4.21	0:44:29
4	2;6.10	1:35:06
5	2;7.14	1:07:33
6	2;8.18	1:04:59
7	2;9.22	0:45:15
8	2;10.20	1:48:21
9	2;11.17	1:27:42
10	3;0.4	1:26:10
11	3;1.5	1:19:42
12	3;2.2	1:56:08
13	3;2.22	1:35:06
14	3;3.2	1:25:36

The transcripts that were chosen to be included in the analysis were selected in order to provide a fairly even distribution of material across the recording period of a year. Exceptions include relevant recordings that were conducted immediately before or after a period of time involving a significant change in Hedda's input. The first of these two events involves Session 9, which was conducted just before the family traveled to South Africa, and Session 10, which was done the same week they returned home to Norway. The second major event was a visit from Hedda's American grandmother, which took place between Sessions 13 and 14, with the recordings being done a day before her grandmother arrived and a day after she returned to the United States. These recordings were included because the shift in input had a significant effect on Hedda's production in English, as we will see later. In addition to these two periods, one other period involving a significant change in Hedda's input took place between Sessions 4 and 5, when Hedda traveled with her parents to the United States for a

family reunion of Hedda's father's family in which she interacted with a variety of English speakers (including other children).

An overview of the Norwegian transcripts included in the analysis is provided in Table 4 below.

Table 4 *An overview of the Norwegian recordings used for the analysis.*

Session number	Age	Recording length
1	2;3.29	0:18:04
2	2;6.5	0:18:05
3	2;6.27	0:15:59
4	2;11.14	0:58:57

In terms of the time frame for the Norwegian recordings, Session 1 roughly corresponds with English Session 2, Norwegian Session 2 roughly corresponds with English Session 4, Norwegian Session 3 falls approximately halfway between English Sessions 4 and 5, and Norwegian Session 4 roughly corresponds with English Session 9. Unfortunately, due to time constraints these are currently the only Norwegian sessions in the Hedda corpus that have been transcribed at the time of writing, and thus the Norwegian transcripts neither span the full year that the English transcripts do nor are they as evenly distributed. The Norwegian sessions included here are also on average much shorter than the English sessions: three of the four Norwegian sessions are only between 15 and 20 minutes long, with only one session (the fourth) being comparable to the English sessions in length. The average length of the Norwegian sessions is 27 minutes and 46 seconds while the average length of the English sessions is 1 hour, 21 minutes and 56 seconds. While this will be remedied as the Hedda corpus continues to be expanded and worked on in the future, it presents a limitation for the analysis conducted in this thesis at this time. Nevertheless, the included Norwegian transcripts provide a useful contrast to the English sessions and demonstrate Hedda's different language production behaviors in Norwegian and English contexts.

For both sets of recordings, utterances were coded as either English or Norwegian, with one language being the default unmarked choice depending on whether it was an English or Norwegian session, and the opposite language being marked (indicated by the presence of [- nor] or [- eng] at the beginning of the utterance in the transcript, with the former being used for the English session transcripts and the latter being used for the Norwegian session

transcripts). Mixed utterances were indicated by the sequence @s at the end of each word in question, indicating that that word was in the opposite language. Words marked with @s in English utterances were Norwegian words, and words marked with @s in Norwegian utterances were English words. Word blends involving morphology from both languages were also marked as mixing using @s. Examples in this thesis will use boldface type to indicate a mixed item instead of the @s sequence for ease of reading.

For the English sessions, there were several instances (particularly single-word utterances) where determining which language the child was using was difficult or ambiguous due to the high number of cognates between Norwegian in English. De Houwer (2009) refers to these types of utterances as "indeterminate utterances." In these cases, vowels and other phonological cues were used to make a decision for the purpose of the transcript, which necessarily codes utterances as having either Norwegian or English as the base language. For an example, Hedda typically pronounced the word *zebra* using the English vowel [i] in the first syllable, while "elephant" was usually coded as Norwegian (*elefant*) due to her stress placement and vowel quality on the final syllable. The most common ambiguous word pair in the transcriptions was Norwegian *ja*/English *yeah*. For this pair, vowel quality was the guide for coding which language the utterance belonged to for the purposes of transcription, but due to the high frequency of this pair of words in the recordings, these were excluded from the analysis.

Hedda's mixed utterances were extracted from the transcripts to form a sub-corpus of data. Each entry was tagged with potentially relevant information, such as the directionality of mixing, whether the utterance was a response or not, who the interlocutor was, which language they were speaking, as well as information about how many words the utterance contained and the number of words belonging to each language. This data was examined to see if any patterns emerged.

3.3 Potential issues/shortcomings with methodology

There are several methodological issues which need to be addressed with regard to this thesis. Fortunately, as part of a larger research project in progress at the University of Tromsø – The Arctic University of Norway, the recordings will continue to be transcribed and the data set used in this thesis will be built upon. As more data is transcribed for the dense corpus, the

possibilities for further research will grow as well. At this point, the number of data points is too few for the sort of quantitative analysis that I would have liked to do, such as determining the potential statistical significance of the changes in MLU observed after periods of increased English input (to provide just one example). As the corpus continues to grow, the potential for statistical analysis will also grow.

Related to the issue of the transcripts is the fact that they are not yet in what could be considered a final form, in the sense that they have not been independently transcribed by separate native speakers of Norwegian and English (as previously mentioned), which would be the ideal scenario. At present, the English transcripts have been transcribed by the author of this thesis (a native English speaker) and proofread by a native Norwegian speaker, both of whom are bilingual in English and Norwegian, and the Norwegian transcripts were carried out by a native speaker of Norwegian but they have not been checked or transcribed by anyone else. This is partly due to the challenge of finding appropriately qualified people who were willing and able to transcribe the files within the time limits of this thesis, since the ideal transcriber for the Hedda corpus is bilingual in Norwegian and English, with one of those two languages being their native language. As mentioned above, however, the research group that this thesis project is affiliated with will continue to transcribe and analyze the data, so these shortcomings will be remedied given more time. Nonetheless, the results of this thesis may be considered somewhat preliminary, given these shortcomings and the fact that further work with the data may provide new insights into Hedda's language development in her two languages. It is my hope that it will nonetheless form a foundation for future work done with the Hedda corpus.

Other shortcomings of this study include the lack of certain measures of Hedda's linguistic development as well as any data regarding absolute levels of input for each language. On the first point, including measures such as CDI forms for both Norwegian and English at different points during the recording period would in theory provide useful information about Hedda's lexical development, but unfortunately no CDI forms have been filled out at the time of this thesis, so we have no information of that sort for the first three years of Hedda's life – only what is present in the recordings. Nonetheless, the recordings do show that Hedda's comprehension is much more developed than her production in her weaker language, English, and while there are many words in English Hedda did not typically produce in her natural speech at certain points, she would be able to produce the English translation equivalents of

Norwegian words when prompted. On the second point – absolute input – there was the intent to use a LENA device to be able to measure absolute input for a sample of weekend days, when both parents were home, but Hedda was exceedingly reluctant to wear the device for extended periods of time (we also tried using the LENA recorder for our shorter weekly sessions, but Hedda took off the vest containing the device herself within ten or fifteen minutes of the session beginning, every time). This makes it difficult to say anything about the absolute input Hedda received in each language on average, but it is possible to say that she without a doubt received proportionally more input in Norwegian than in English.

3.4 Conclusion

This chapter has provided an overview of the methodology for the data collection, transcription, and analysis for this thesis. While the limited number of recordings currently transcribed restricts the scope of analysis this thesis can cover, the Hedda corpus will fortunately continue to grow and become a rich resource for BFLA research.

4 Analysis

This chapter will provide an overview and analysis of the data collected in the 14 English transcripts and the 4 Norwegian transcripts. First, the assertion of Hedda's Norwegian dominance will be addressed through a number of different measures, all of which will also provide a fuller picture of Hedda's language production and development in general. The focus will be on her language choice and code mixing in her production, with an eye toward the relevant factors that influence these behaviors. After addressing her Norwegian dominance, Hedda's language mixing data will be examined more closely. Her case will be compared to other case studies of bilingual children's code mixing, as some of her behaviors can be predicted by existing research while others appear to be more unusual.

At the end of the chapter, Session 14 will be given extra attention, as it stands out from the others in terms of Hedda's language production. Much of what we can say about Hedda's English language development and production stems from what she produced during this session.

4.1 Hedda's language dominance

While Hedda's dominance in Norwegian would be obvious to any English speakers interacting with her, establishing this dominance on the basis of the data is nevertheless a necessary step. There are a variety of ways to measure dominance in bilingual children, and often researchers use a combination of measures. Paradis et al. (2003) looked at children bilingual in French and English in Canada, and classified children as dominant in one of their languages if they demonstrated higher performance in one language on three out of five measures.

I will use three measures to demonstrate Hedda's dominance, but as will become clear, for the majority of Hedda's English sessions, she demonstrates a marked dominance in Norwegian for all three measures. When the input situation changed, however, the balance of Hedda's dominance shifted. My focus is on measures that use Hedda's naturalistic language production, since, as Lanza (1997) has noted, many of the measures used to determine dominance in adult bilinguals are carried out in experimental or test settings and are not suitable for child bilinguals at the age of 2 or 3.

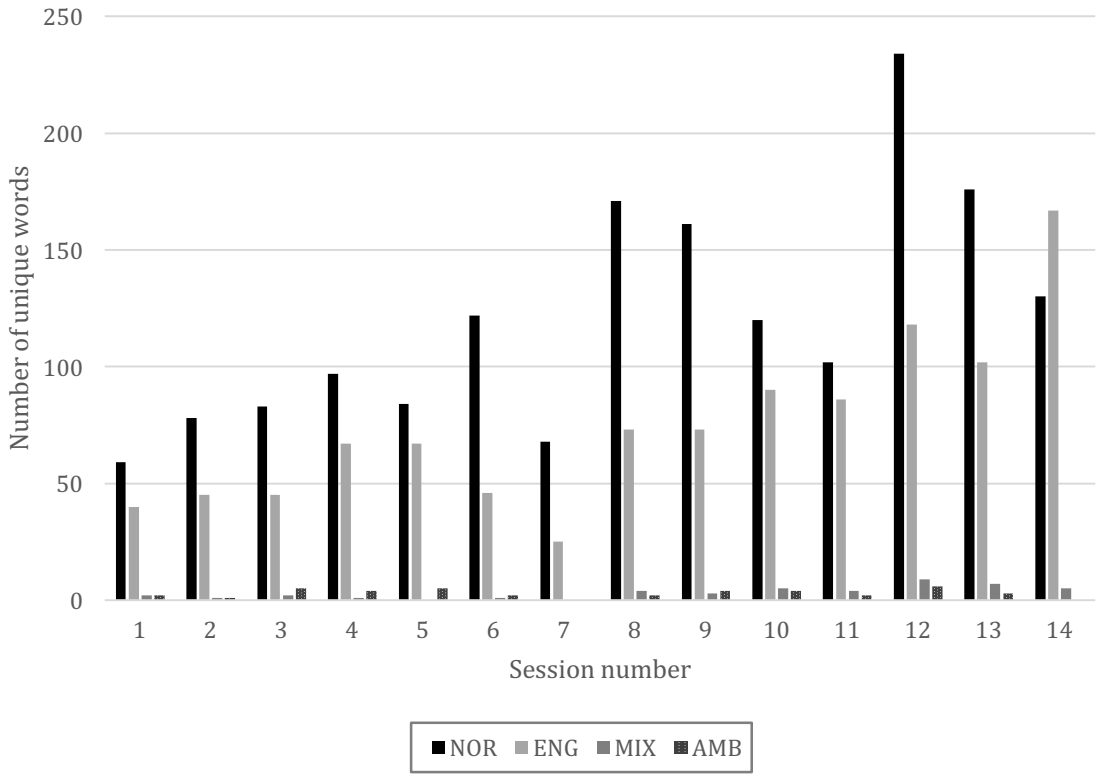
Section 4.1.1 will provide an overview of the number of unique words used by Hedda in each language during both the English and the Norwegian sessions, which will demonstrate the development of her productive vocabulary over time and tell us something about language choice in the two different contexts. Subsequently, Section 4.1.2 will examine the overall mean length of utterance for Hedda's sessions. Section 4.1.3 follows with a summary of the proportion of utterances by language for each session, and finally, there will be a word about the directionality of mixing in section 4.1.4, although I am not using this measure as an indicator of Hedda's dominance. Instead, I will show that having been established by the first three measures, Hedda's dominance can be used to predict the directionality of mixing in her mixed utterances.

4.1.1 Unique number of words in each language

The FREQ program in CLAN was used to generate a list of types and tokens for each transcript. Using this list, types (that is, unique words) were sorted into one of four groups: Norwegian ("NOR"), English ("ENG"), Mixed ("MIX"), or ambiguous ("AMB"). For the purpose of determining the number of unique words Hedda used in each language during recording sessions, words that are derived from one another through verb conjugation or derivational morphology (as in English *make/making* and *teapot/teapots* or Norwegian *rygg/ryggen*) were counted as one unique word, rather than separate types. Names of people, words referring to Hedda's parents (whether *mamma* or *mommy*), and *yes/yeah/no* (and their Norwegian counterparts *ja/nei*) were excluded from these counts. Items in the Mixed category consist of single words that show a blend of both Norwegian and English morphology, as in the verb *eat-e*, which uses the English verb "eat" with Norwegian infinitive ending *-e*. Items in the Ambiguous category consist of short words that could be assigned to either language and are difficult to assign a language value to even in context, such as *ball*.

The chart in Figure 1 provides an overview of the number of unique words in each language during Hedda's English recording sessions. The absolute number of unique words was used in this case because it sheds more light on Hedda's lexical diversity in general than would using proportions alone. One of the things that becomes immediately visible is that Mixed and Ambiguous categories were relatively minuscule for all sessions, never including more than 10 types (and typically including only 1 or 2, and sometimes zero).

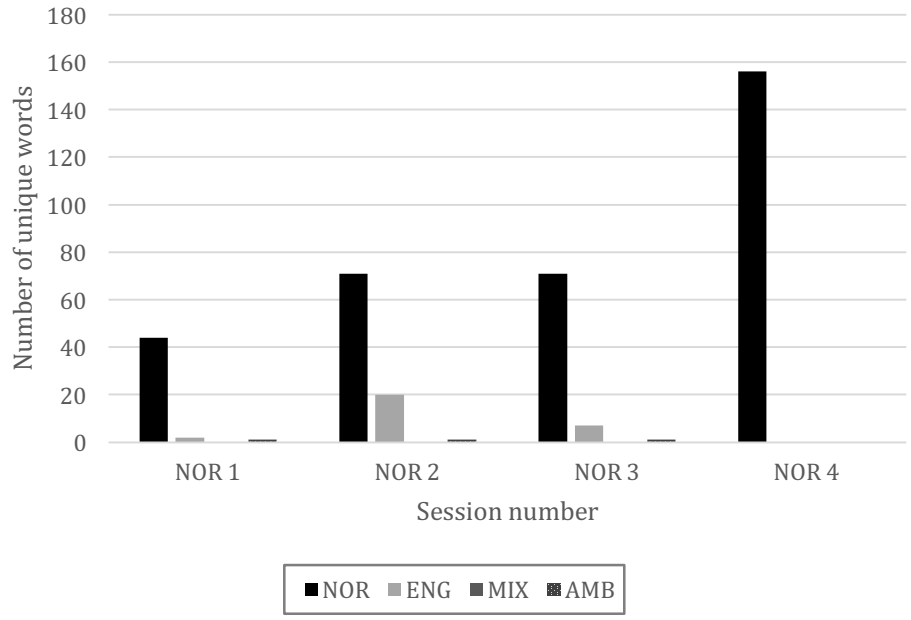
Figure 1 *Unique number of words in each language for English recording sessions*



A general upward trend can be seen over the course of the year between Session 1 and Session 14, indicating that Hedda's productive vocabulary is growing over time. The interesting thing about comparing the number of unique words in English and Norwegian is that in all but one session, Hedda used a larger variety of Norwegian words than English words. This is true even in the case of one of the two sessions where she has a larger number of English utterances than Norwegian ones, Session 10 (see section 4.1.3 below). This very clearly suggests Norwegian dominance, because Hedda appears to produce a larger variety of Norwegian words than English words, even when addressing English speakers. The one exception, recording session 14, took place the day after her grandmother's weeklong visit, and as we will see as we look at the data in other ways, this session marked a turning point in Hedda's language production and dominance – because dominance is of course not static and can change depending on a number of factors.

The same process was carried out for the four Norwegian transcripts. The corresponding chart can be seen below in Figure 2.

Figure 2 *Unique number of words in each language for Norwegian recording sessions*



While the small number of Norwegian sessions and their shorter length does not make for an ideal direct comparison between the Norwegian and the English sessions, it does mean that these four sessions were able to be examined in greater depth than the set of English transcripts. This reveals that the data within the chart alone could be misleading. Data encoded in the transcript that the FREQ program in CLAN ignores, as well as contextual clues from the content of the utterances, suggest that the values for unique English words cannot be taken at face value. These factors vary for each session, so I will provide a quick overview for each below.

Norwegian Session 1 contains two unique English words amounting to approximately 4.2% of unique words based on the list generated by CLAN, but a closer inspection of the transcript itself reveals that both of these words were marked with a [?] by the transcriber, meaning that she was unsure of exactly what Hedda was trying to say and she opted to transcribe her best guess. While Hedda's father entered the room at the very end of the recording and spoke a total of five utterances, these two potential English words spoken by Hedda occur in the first half of the recording, when only Hedda and her mother are interacting. The reliability of this value is thus uncertain.

Norwegian Session 2 contains a much larger proportion of unique English words at first glance (20 out of 92 total unique words, or approximately 21.7%), but in this case context

must be taken into account. With the exception of only one unique word, Hedda's English words in this session were either part of an English language song the family was singing together, or she was making direct reference to the song. Hedda's father was present for more of this session, and the whole family (i.e. Hedda and both her parents) participated in singing both of the English songs. There is other evidence in the corpus of topic influencing Hedda's language choice, so her use of English when referring to English songs is not surprising, nor is the fact that the family is singing songs in English when her father is present. For the remainder of the recording, Hedda uses exclusively Norwegian with her mother with the exception of one word.

Norwegian Session 3 contains fewer unique English words again, but still more than Session 1 (accounting for approximately 8.9% of unique words). Out of 7 unique English words, 5 belong to utterances marked with [?] indicating that the transcriber was unsure of what Hedda said and opted to make a best guess, as with the first Norwegian session. The other two unique English words belong to a mixed utterance Hedda used frequently with both her parents and the investigator throughout the recording period: *æ kan ikke **see you**!* ('I can't see you!'). This phrase appears to be restricted to a single context, in which Hedda is playing a peekaboo game with one or more of the adults. As such, we cannot be entirely sure that Hedda has parsed *see you* as an English phrase. Once again, there are no clear cut examples that would suggest Hedda is mixing the amount of English into her Norwegian suggested by the numerical value in Figure 2.

Norwegian Session 4, the longest session and the only one comparable to the English sessions in length, contains no English words. Almost the entire session involves only Hedda and her mother, as her father is working outside. At the very end, he comes back in, effectively bringing the session to an end. He speaks a total of five utterances in English to Hedda, but she does not actually respond to any of them (though she laughs). As this is currently the only session featuring such a long uninterrupted period of time with Hedda and her mother, we cannot say much definitively about whether this session represents an exception or the rule, but based on the observations of Hedda's parents, as well as the caveats for the English present in the other sessions, we may reasonably assume that this session represents Hedda's typical language behavior with her mother and in other Norwegian situations.

One thing that becomes very clear immediately is that Hedda's language production is not the same across the board: while the English recordings contain a large amount of Norwegian, the Norwegian recordings contain relatively little English, and upon closer inspection much of the English that is there can either be explained through context or could be considered unreliable due to the transcriber employing the best guess strategy. Some of this very likely goes the other way, in that some of the Norwegian used during the English sessions is present because Norwegians songs or stories are being discussed, Hedda's mother is often present and even speaks Norwegian in addition to English during these sessions, or other context-specific reasons. Nevertheless, the investigator is Hedda's interlocutor for a majority of the time during the English sessions, and Hedda often speaks Norwegian to her as well. The complete lack of word blends (the "MIX" category) in the Norwegian sessions is a relevant factor given their presence in the English sessions. Additionally, the fact that unique Norwegian words outnumber unique English words in all but one session for the English sessions is striking, and when the eighteen sessions are taken together they certainly indicate that Hedda is dominant in Norwegian, with a larger productive vocabulary in Norwegian than in English.

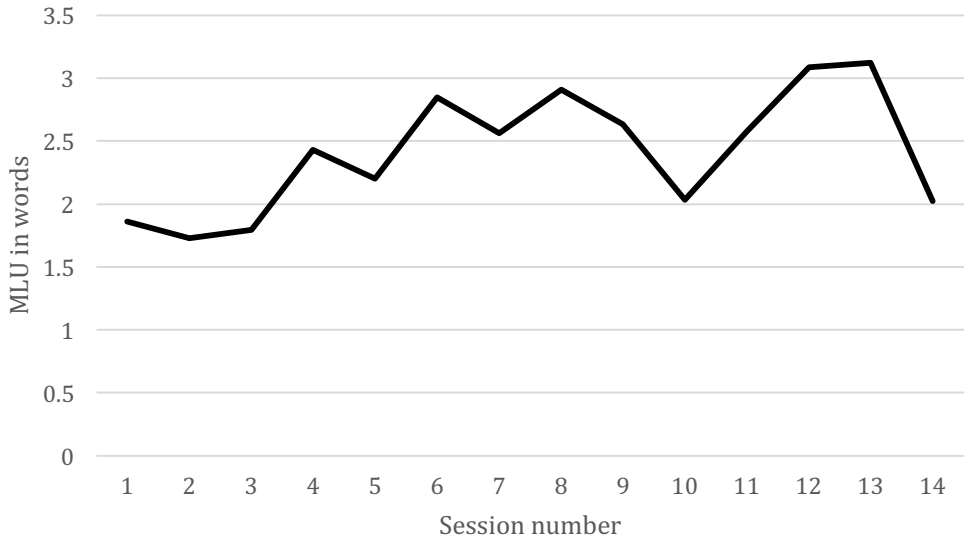
4.1.2 Mean length of utterance

Using CLAN, mean length of utterance (MLU) was calculated for each speaker. As the transcripts were not coded for morphology at the time of writing, words were used as the unit for MLU. An overview of these values can be seen in Table 5 below. Hedda's MLU ranges from 1.729 to 3.122 and shows an expected upward trajectory throughout the year, with an overall rising trend as her sentences become longer and more complex. Figure 3, however, demonstrates several noticeable low points within that upward trajectory: notably, Sessions 5, 10, and 14. All three of these recordings were the first sessions following a period of increased English input, and the latter two coincide with a higher proportion of English utterances per recording (we will come back to this point later). One other point that is obvious based on Table 5 is that the father was not present in the home during Sessions 6 or 11. This is relevant given that he is Hedda's primary source of regular English input.

Table 5 *MLU for each speaker in the main dataset of English recordings.*

Recording session	Child	Investigator	Mother	Father
1	1.861	3.795	4.091	4.536
2	1.729	3.844	4.938	5.2
3	1.794	4.387	4.5	4.411
4	2.433	4.079	5.525	6.084
5	2.202	4.13	4.899	4.962
6	2.849	4.541	4.738	NA
7	2.561	4.519	4.088	5.5
8	2.91	4.299	5.262	5.446
9	2.636	4.132	5.226	5.338
10	2.022	4.342	4.924	5.613
11	2.581	3.59	6.769	NA
12	3.09	4.003	4.28	5.323
13	3.122	4.27	6.405	6.674
14	2.022	2.865	5.478	7.24

Figure 3 *Child's MLU for English sessions throughout the recording period.*



While some theoretical issues have been raised with regard to using MLU as a measure of language development (particularly MLU based on the number of words, rather than morphemes), it is nevertheless a useful measure to examine with Hedda's changes in input in

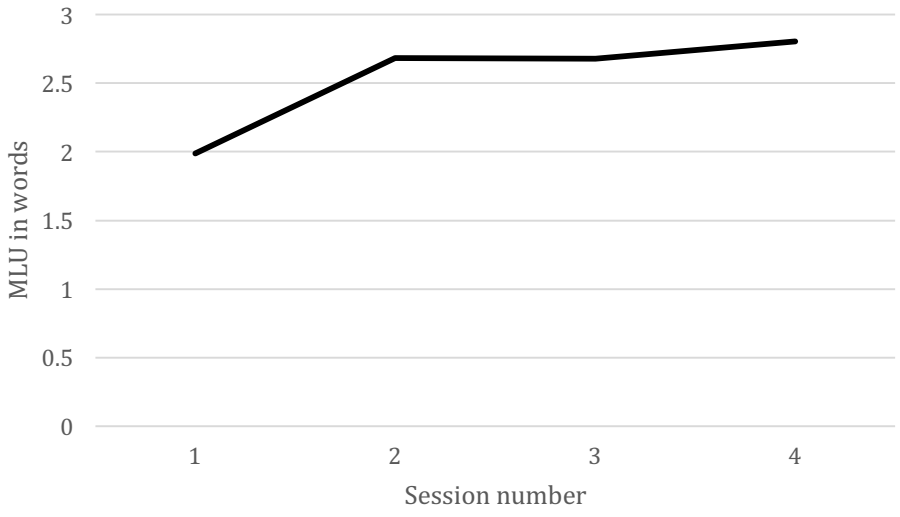
mind. One thing that is noteworthy is that the latter two sessions in which we see a lower MLU, Sessions 10 and 14, are also the only two sessions in which Hedda's English utterances outnumbered her Norwegian utterances. One possible explanation for this might be that Hedda's MLU in English is on average shorter than in Norwegian, but MLU will have to be calculated for each language separately before that can be definitively determined.

The same process was carried out for the four Norwegian transcripts. While these span a shorter period of time, a similar upward trend is observed in general. Unfortunately, the Norwegian transcripts corresponding to the periods immediately after an increase in English input have not yet been completed, but this work will be carried out in the near future. What can be said is that the numerical value for MLU is slightly higher in the Norwegian sessions than for the corresponding English sessions in all four cases.

Table 6 *MLU for each speaker in the Norwegian recordings.*

Recording session	Child	Mother	Father
1	1.986	5.209	2.4
2	2.684	5.19	7.6
3	2.681	4.55	4.75
4	2.803	5.328	6

Figure 4 *Child's MLU for Norwegian sessions throughout the recording period.*



While it is difficult to glean very much information from the limited Norwegian session data currently available, it is notable that the values for MLU are all slightly higher for the Norwegian sessions than they are for the corresponding English sessions. While it seems unlikely that this difference is statistically significant, we may hypothesize that Hedda's average MLU when speaking in a Norwegian context is generally higher than her average MLU when speaking in an English or bilingual context, meaning that her Norwegian utterances are likely longer than her English ones on average. This is a very tentative hypothesis, however, since the English sessions do contain so much Norwegian. Calculating the MLU values for each language in each session would be the best way to test this, but this task can not be accurately calculated automatically using CLAN and must be done manually, thus it was outside the bounds of the timeframe of this thesis. These values will be calculated for future work involving the Hedda corpus, however, in order to address this question.

4.1.3 Proportion of utterances in each language

Given the amount of Norwegian present in the English recordings, a breakdown of the proportion of utterances in each language per recording is a useful tool in general, not just for establishing Hedda's dominance but for obtaining a larger picture of her behavior in language production in an English-speaking setting.

Each transcript was examined manually, with each of Hedda's utterances being assigned to a category: Norwegian ("NOR"), English ("ENG"), or Mixed ("MIX"). Utterances containing only the name of a person or character, words referring to Hedda's parents, or onomatopoeic sounds were excluded. Also excluded were longer utterances consisting of only one language with an unknown item (indicated by an "xxx" in the transcript) present, as it is impossible to say whether these utterances were monolingual or mixed. Utterances containing an unknown item "xxx" were included in the analysis if they also contained both Norwegian and English morphemes, because regardless of which category the unknown item belongs to, it is possible to say without a doubt that it is a mixed utterance. The last group of excluded utterances included those consisting of just *yeah* or *ja*, due to the phonetic similarity of these two words. Other one-word utterances, including *yes*, *no*, and *nei*, were included. Whether the inclusion of these one-word utterances is good practice for gaining a picture of Hedda's overall language production is perhaps a matter of debate, but given the high number of one-word utterances present in her transcripts, I wished to include them.

Figure 5 *Proportion of utterances for each language during English sessions*

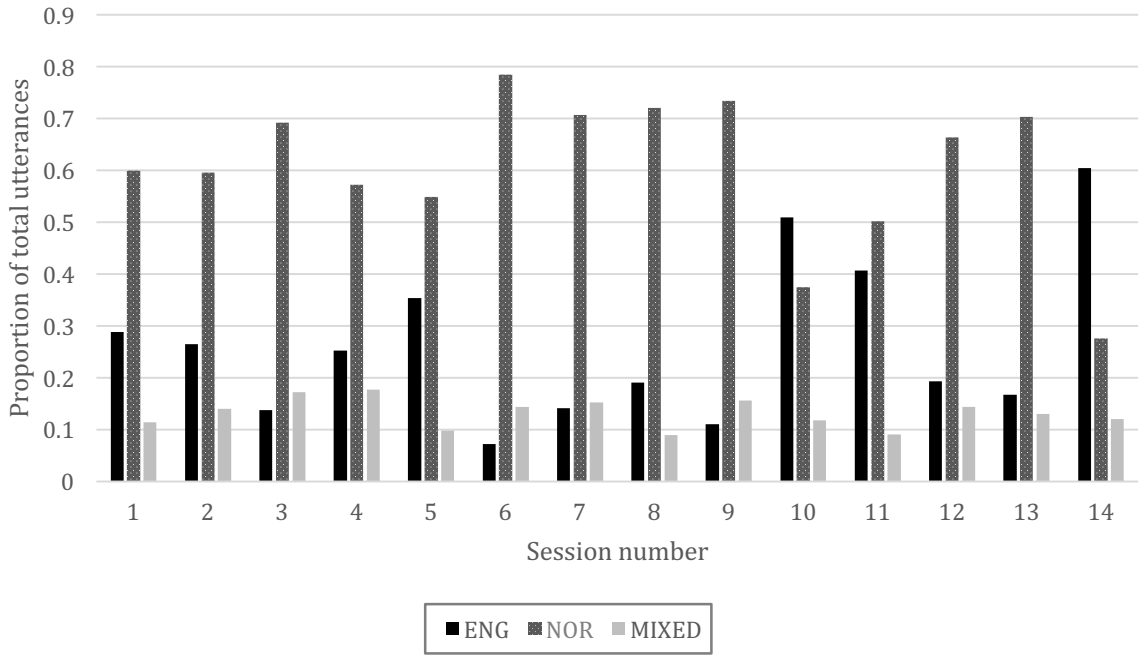
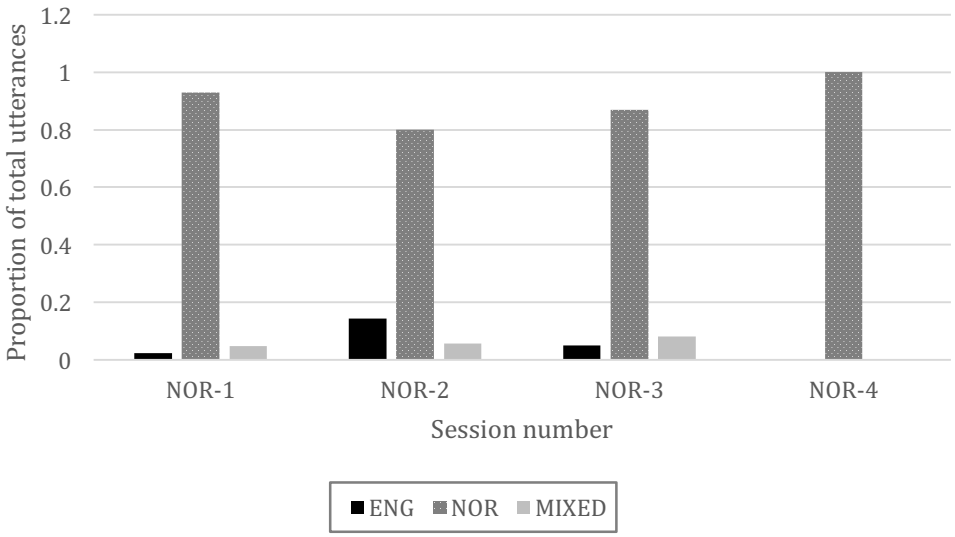


Figure 5 presents an overview of the proportion of utterances in each category for Hedda's English sessions. The absolute number of utterances has been converted to a fraction of 1 in this case in order to compare the amount of English versus Norwegian being produced by Hedda during each session. Fully Norwegian utterances greatly outnumber fully English utterances for a majority of the sessions, but two sessions contain a larger number of fully English utterances. When compared with Hedda's overall MLU as seen in Figure 3, it becomes clear that these are the same sessions in which Hedda's MLU sees a noticeable dip - that is, the two sessions immediately after two of Hedda's periods of increased English input. Session 14 sees the most sizable shift, and this also coincides with the only time Hedda uses a larger variety of English words than Norwegian words during an English session. Not only that, but the effect of this last period of increased English input during the visit from Hedda's grandmother seems to have had a sustained effect – while the following recordings have yet to be transcribed, both the investigator and Hedda's parents have commented on the sustained increase in English production, noting that Hedda often speaks English to her father now where she typically used mostly Norwegian before.

In order to drive home the argument for Hedda's dominance in Norwegian, however, it is helpful to look at the same proportion of utterance counts for the four Norwegian transcripts.

Figure 6 *Proportion of utterances for each language during Norwegian sessions*



As we can see in Figure 6, Norwegian utterances positively dominate here. Once again the caveats brought up in section 4.1.1 apply: the larger proportion of English utterances present in Norwegian Session 2 can be accounted for by the English songs the family sang together, and many of the words marked as English words were the transcriber's best guesses, meaning they may not actually be English. While Table X containing the MLU counts for the Norwegian sessions indicates that the father had a total of three utterances during the final Norwegian session (NOR-59), these occurred at the very end of the recording session and did not elicit any speech from Hedda, only laughter. This fourth session is the only Norwegian session comparable in length to the English sessions, being around an hour. Remarkably, this session contained 100% Norwegian utterances, with no English or mixed utterances at all.

One clear implication of these two charts is that Hedda can very clearly distinguish between Norwegian contexts and English contexts (or perhaps "bilingual contexts" would be a more accurate description to use here) in that she behaves like a monolingual Norwegian child in monolingual Norwegian contexts, and she produces English in contexts where other English speakers are present. This is very strong evidence that she separates her languages and has done so since the beginning of the recording period covered in this thesis. While Hedda's specific behavior is different than Kate's from De Houwer (1990), the fact that her language usage looks different in different contexts echoes De Houwer's findings regarding Kate's separation of her languages. For Hedda, the huge presence of Norwegian production in conversation with English speakers points to her very strong dominance in Norwegian for the majority of the recording period. Notably, mixed utterances persist throughout all fourteen

English sessions, even once the English starts to gain some ground on the dominance scale and even though mixed utterances of the type she produces are very nearly absent in her input. As future work is done on the Hedda corpus, it will be interesting to see whether the proportion of mixed utterances is sustained in the months following Session 14 or whether there is an increase or decrease for some period of time.

4.1.4 Directionality of mixing

Lanza (1992, 1997) used directionality of mixing as a measure of dominance for her subject Siri. She pointed at several generalizations that could be made about Siri's mixed utterances as an indicator of this dominance. Namely, while English lexical morphemes could co-occur with both Norwegian and English grammatical morphemes, the reverse was largely untrue for Siri: Norwegian lexical morphemes could only co-occur with Norwegian grammatical morphemes, so while blends like *look-e* (English *look* + Norwegian verbal *-e* suffix) were possible, blends such as *husk-s* (Norwegian *husk* + English verbal *-s* suffix) were not. While there were exceptions to these general patterns, Lanza argued that "[t]hese co-occurrence constraints indicate the prevalence of a Norwegian grammatical framework in Siri's speech." Based on Hedda's strong Norwegian dominance indicated by the other measures, we can make an educated guess at what her code mixing will look like. A pattern like Siri's would be very much expected, so that both functional and lexical morphemes can be mixed from Norwegian into English, but functional English morphemes should not appear in Hedda's Norwegian utterances. The results of the analysis are provided in the next section.

4.2 Code mixing

Because Hedda produced so many mixed utterances during her "English" recordings, a systematic analysis of her mixing within utterances seemed like it might be able tell us something interesting. The mixed utterances Hedda produced during the 14 English sessions were extracted and compiled in a single database. Each utterance was tagged for a number of different pieces of information: the length in words, the "base language," the type of mix (more on how this was defined below), and whether or not Hedda was repeating a word or phrase just used by her interlocutor, to name a few. This system of categorization works fairly well in some ways but not necessarily in others – in particular, determining the base language

and the type of mix for each utterance proves especially challenging for two-word utterances, of which there are many in the database.

There are a number of different ways that base language could be defined or determined depending on which criteria you consider to be most important. Given that in these recording sessions Hedda was almost always speaking to an interlocutor that used English with her (most often the investigator), one might consider the base language to be English due to the English context (or at the very least, the goal of an English context). Hedda produces so much more Norwegian than English for the majority of the recordings, however, that it does not make sense to consider English the base language for all of her mixed utterances. Hedda was accustomed to largely dilingual interactions with her father, for instance, in which he spoke English to her and she replied in Norwegian. This comes back to the fact that an "English context" for Hedda, and consequently a monolingual English mode, are almost nonexistent, while a bilingual Norwegian/English context invoking a bilingual mode with both Norwegian and English activated was typical for her home environment. Thus, it is not only possible but likely that the base language for many of Hedda's mixed utterances is Norwegian. Syntax can be another cue to determine the base language, particularly in utterances where only one word is mixed (and there are many examples of this), but this again can be problematic in two-word utterances.

The results for the four mix categories are included below. While acknowledging the potential circularity of the relationship between language dominance and the directionality of mixing (does dominance in one language drive the directionality of mixing or does the directionality of mixing indicate language dominance?), we can use the other measures of dominance included in section 4.1 to establish Hedda's Norwegian dominance, and then make predictions about the directionality of mixing we will see in her mixed utterances based on that.

Given her Norwegian dominance established by the other three measures, we would expect to see Norwegian functors in utterances where English can be considered the base language, and any mixing in utterances where Norwegian is the base language would be more likely to be lexical mixes, not functors. Lexical Norwegian mixes in English utterances would also not be considered unusual, but English functors being mixed in sentences where Norwegian is the

base language would certainly be unexpected, following the pattern that Siri's language mixing showed (Lanza 1992).

The four categorizations for Hedda's mixes are broken down by base language and basic type:

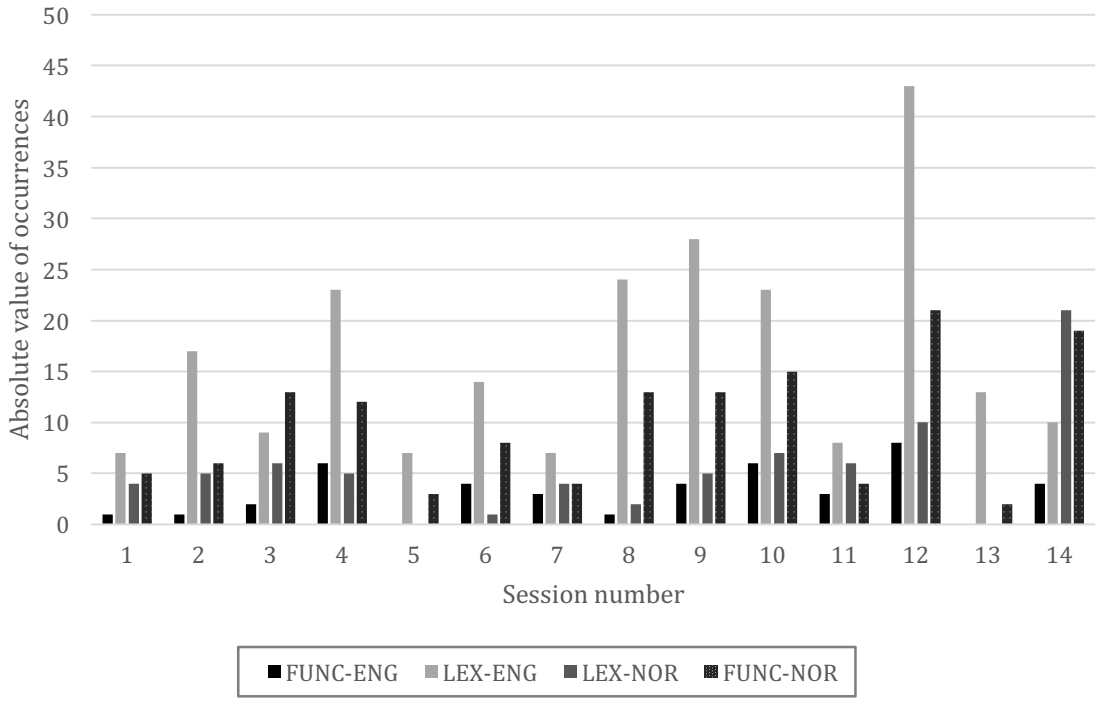
I. Norwegian as the base language
(a) FUNC-ENG (English functional items)
(b) LEX-ENG (English lexical items)

II. English as the base language
(a) FUNC-NOR (Norwegian functional items)
(b) LEX-NOR (Norwegian lexical items)

Some utterances contained multiple mixes and thus more than one type; both categories were counted for those utterances. In cases of consecutive identical repetitions of a mixed utterance, only the first utterance was counted. A summary of the frequency of each category for the 14 sessions is shown in Figure 7.

Based on Lanza's (1992, 1997) work, we can make certain predictions about the frequency of the different mix types. Given that the English sessions were designed to evoke an English context, we would expect a relatively high rate of FUNC-NOR in her mixed utterances (where the base language is English). LEX-NOR is also expected to be present in English utterances. Since I have established that the number of Norwegian utterances almost always outweighs the number of English utterances, however, we can also predict that LEX-ENG will be relatively high in her mixed utterances where Norwegian is the base language, but that there will not be a high number of FUNC-ENG mixes, because following Lanza, the latter would be an indicator of English dominance (or at the very least, no discernable dominance, if mixing of functors went both ways). Hedda's actual results are displayed in Figure 7 below.

Figure 7 *Type of mixing in Hedda's mixed utterances (absolute values)*



The expected pattern does play out over a majority of the recording period. Hedda's most frequent mix types by far are Norwegian functors in English utterances (FUNC-NOR) and lexical English items in Norwegian utterances (LEX-ENG). LEX-ENG is almost always the most common type of mixing, in fact. Given that Hedda produces more Norwegian utterances than English utterances in all but two of the recordings, this is not surprising. English functors mixed into Norwegian utterances are predictably lower than other mix types, which does support Lanza's assertion that the directionality of mixing is related to dominance. It is also interesting to note that the picture changes for Session 14, which is the outlier in regards to Hedda's dominance measures as well. In Session 14 where we have more English utterances than Norwegian utterances *and* a greater variety of productive English vocabulary than Norwegian vocabulary, the balance shifts with regard to the type of mixing Hedda is producing.

The absolute values in Figure 7 were converted to percentages for each session. These can be seen in the table below.

Table 7 *Percentage of mix types for English sessions*

SESSION:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
FUNC-ENG	5.88	3.45	6.67	13.04	0	14.82	16.67	2.5	8	11.77	14.29	9.76	0	7.41
LEX-ENG	41.18	58.62	30	50	70	51.85	38.89	60	56	45.1	38.1	52.44	86.67	18.52
LEX-NOR	23.53	17.24	20	10.87	0	3.70	22.22	5	10	13.73	28.57	12.2	0	38.89
FUNC-NOR	29.41	20.69	43.33	26.09	30	29.63	22.22	32.5	26	29.41	19.05	25.61	13.33	35.19

It should be mentioned that this method of categorizing mixed elements by type works better for some sessions than for others, as Hedda's mixed utterances unsurprisingly become more complex over time. Initially, many of Hedda's mixed utterances contain only one word or element from the opposite language. By the last few sessions, there are several examples of mixed utterances featuring self-corrections when Hedda realizes mid-utterance that she should be using the other language, as in examples (9)-(10) below.

(9) *CHI: <**du må l(ese)>** [///] can you read this? (Session 12, 3;2.2)

In this example, Hedda has cut herself off before completing the word *lese* ('read') and switched to English to make the entire request again.

(10) (Session 12, 3;2.2)
*CHI: [- nor] den (.) må du lage en liten båt.
that you must make a little boat
'You have to make a little boat.'
*INV: what?
*CHI: [- nor] det må du lage et stort båt.
it must you make a big boat
'You have to make a big boat out of it.'
*INV: you have to tell me in English.
*CHI: [- nor] du må lage et (.) **a boat**.
you must make a
'You have to make a boat.'

Again, in the last utterance Hedda begins her utterance in Norwegian, then stops herself and switches to English to complete the utterance. This time she merely completes the utterance in English rather than repeating the whole thing. In addition to examples such as these, Hedda shows more and more mixing at the phrasal level that looks more like adult code mixing, as in the following examples:

(11) *CHI: det er ikke **the face**. (Session 10, 3:0.4)
it is not
'It's not the face.'

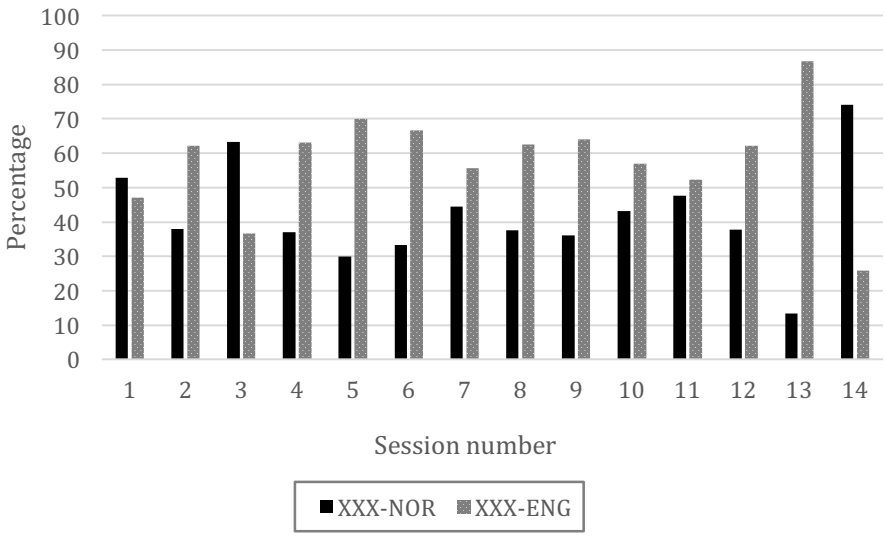
(12) *CHI: **æ tror det** look like green. (Session 12, 3;2.2)
I think that
'I think that looks like green.'

Mid-utterance switches such as these are more difficult to categorize using the four-category system due to their very nature. Nonetheless, these categories can tell us something useful about Hedda's mixing habits, and a further breakdown and comparison of the categories may prove useful.

4.2.1 Breakdown of mix types

The mixed utterance types were grouped by base language, so that the percentages of LEX-NOR and FUNC-NOR mixes were combined, as were the percentages of LEX-ENG and FUNC-ENG.

Figure 8 *Percentage of mix type by base language*

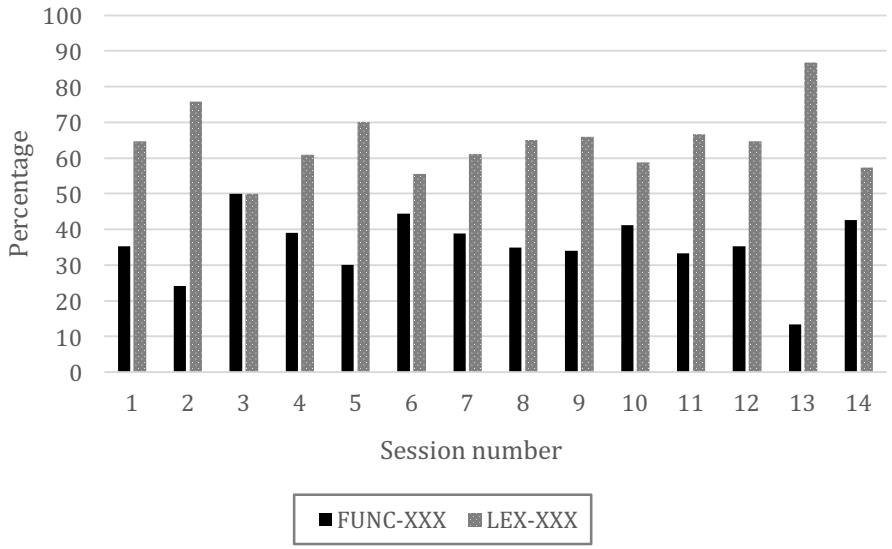


In Figure 8, XXX-NOR represents Norwegian elements mixed into utterances where English is the base language, and XXX-ENG represents English elements mixed into utterances where Norwegian is the base language. Hedda mixes English words into Norwegian utterances more often than the reverse for 11 out of 14 sessions. Given that she produces more Norwegian

utterances, this is unsurprising. Session 14 bucks the trend with a substantially higher proportion of Norwegian words mixed into English than any other session, but this may be a result of the proportionately higher number of English utterances in this session.

Figure 9 provides a comparison between Hedda's mixing of functors and of content words (regardless of base language).

Figure 9 *Percentage of mixing of functors vs. content words*



FUNC-XXX represents functors mixed into the opposite language while LEX-XXX represents the content word mixes. On the whole, Hedda uses lexical mixing more than she does mixing of functors (with Session 3 seeing an equal number of both types). But does this breakdown look the same for each base language? An overview is presented below in Figures 10-11.

Figure 10 *Mix types when base language is English*

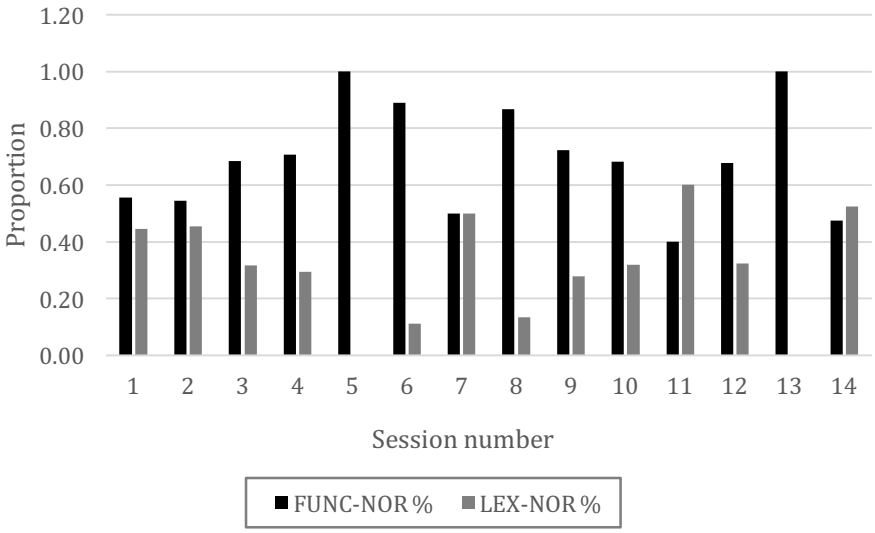
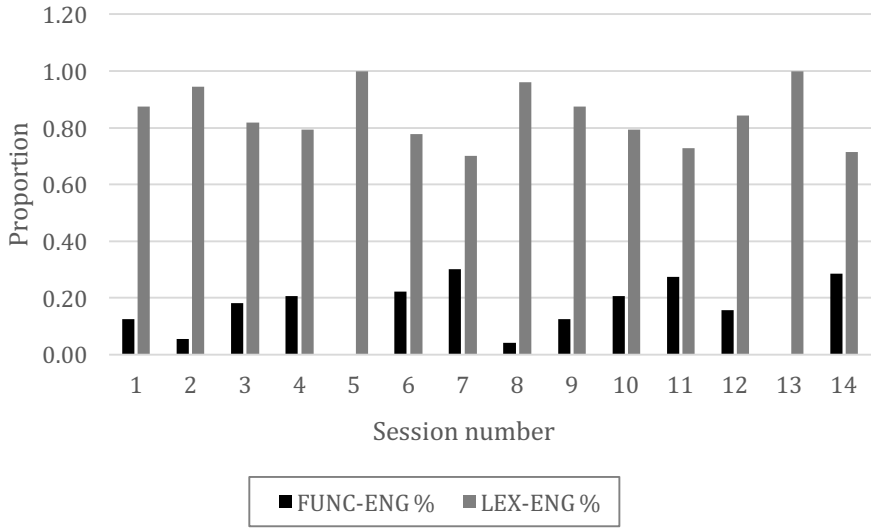


Figure 11 *Mix types when base language is Norwegian*



The answer is a very clear no: the overall pattern disperses and the numbers look quite different for each base language. When the base language is English, Hedda mixes more Norwegian functors than lexical words into her utterances, but the reverse is true when the base language is Norwegian: English functors are rarely mixed into Hedda's utterances. All of this is unsurprising given her Norwegian dominance.

While it would be interesting to try and compare Hedda's use of different mixing types with different speakers, a majority (approximately 75%) of Hedda's mixed utterances during the English sessions were directed at the investigator. Future work with the Hedda corpus may be

able to explore this question, however, as the English sessions involving only Hedda and her father will be transcribed and can be compared to the sessions with the investigator present to see if there are any differences. It would be particularly interesting to compare any results with the findings of De Houwer (1990) and Lanza (1992, 1997), who both found that the interlocutor influenced mixing habits for their subjects.

4.2.2 Mixing as a gap-filling strategy

Existing studies of BFLA have shown that mixing in child language can be a lexical gap-filling strategy, with Greene et al. (2012) stating that bilingual preschoolers "demonstrate the use of code-mixing as a compensatory strategy to fill lexical gaps" and Genesee (2008) arguing that "bilingual children who code-mix to fill lexical gaps are overextending the same way as monolingual children, except that they draw on the lexical resources of two languages." While much of Hedda's code mixing cannot be explained this way, as she uses Norwegian words in English contexts when she knows the English equivalents (and can produce them on command), some of her mixing does seem to be motivated by trying to express herself and being unable to find English words to do so. I provide an example in (13) where Hedda clearly knew the English word, but used the Norwegian instead (perhaps because it was more accessible), and then I will move on to a discussion of mixing that appears to be motivated by the need to fill a gap in her English production.

(13) Hedda and the investigator are putting a sweater on one of her dolls.
*CHI: [- nor] knappa! (Session 9, 2;11.17)
buttons!
'Buttons!'
*CHI: [- nor] ja (.) knapper.
yes buttons
'Yes, buttons.'
*INV: oh (.) oh those.
*INV: do you know what those are called in English?
*CHI: yeah.
*INV: what are they called?
*CHI: buttons!

As indicated by the data, Hedda used much more English during Session 14, which involved an extended game of pretend, but she struggled to form narratives in English and often resorted to mixing or simply using Norwegian in these instances when she became frustrated. Her difficulty accessing the English she needed was apparent in frequent pauses, hesitations,

or even a rising intonation making statements sound like questions. In the following example, Hedda and the investigator are discussing the bedroom door and a troll Hedda pretended to see.

(14) (Session 14, 3;3.2)
*INV: what's happening?
*CHI: the door's (.) **helt** [?] (..) the door's have been (.) **en farge** xxx **til en** (.) a door!
completely a color to a
*INV: a door.
*CHI: that's a door!
*INV: did the troll go away?
*CHI: yes.
*INV: where did it go?
*CHI: it's (..) xxx that (…) **det hoppet i do**!
it hopped in toilet
'It's…that…it hopped in the toilet!'
*INV: in there?
*CHI: yeah.
*INV: <did the> [/] did the troll go in the toilet?

In the example above, 'xxx' indicates an unintelligible item or word, and the sequence [?] following a word indicates that the transcriber is uncertain if this is the word being used but they have provided their best guess. The [/] symbol indicates that the material preceding it in brackets is being repeated. Hedda is clearly struggling to find the words needed to explain what she is imagining in English, and in the sixth line, after a long pause she finally gives up and says it in Norwegian instead, *det hoppet i do!* ('it hopped in the toilet!'), pointing for emphasis. Unlike cases where Hedda uses Norwegian even when she knows the English equivalents, these are cases where she made a conscious effort to speak English and found herself with gaps. It will be interesting to see in future work on the Hedda corpus how much of her mixing takes on this form if her increase in English production is sustained.

4.3 Discourse strategies

While the data collected for this thesis is of a naturalistic type and thus is not as rigidly structured as data collected during experiments would be, there was nonetheless attention paid to discourse strategies used with Hedda during the sessions and their effects on her language choice. The findings discussed in this section will largely be general trends drawn from observation and an examination of the transcripts, as it is difficult if not impossible to say anything definitive about Hedda's language choice with regard to discourse strategies

employed by her interlocutors as it would be difficult to control for the various other factors. Nonetheless, I hope my discussion of how Hedda responded to different discourse strategies at different points throughout the recording period provides some insight into how discourse strategies may influence the bilingual child's language choices.

While I have previously discussed and defined different discourse strategies earlier in this thesis, a brief reminder of the strategies I am interested in here is useful. The relevant discourse strategies include the **minimal grasp strategy**, the **expressed guess strategy**, and the **repetition strategy**, which are listed here in order from most effective to least effective when it comes to establishing a monolingual context for the bilingual child. Additionally, the **move on strategy** should be mentioned as well, falling at the end of the spectrum after the repetition strategy, being the least effective at inducing a monolingual context. In the move on strategy, the parent or interlocutor simply accepts the child's utterance in the "wrong" language (I put wrong in quotes here because in a bilingual context between two bilingual speakers, there is no real reason to consider either language the wrong language and a monolingual context may not be the parent's goal) and carries on with the conversation, having understood the child's utterance and accepted it. The speaker who employs the move on strategy is likely prioritizing communicating effectively with the young child over trying to create a specific monolingual language context that would encourage the child to use one language and not the other.

I have mentioned previously in this thesis that Hedda's father typically employed either the repetition strategy or the move on strategy when speaking with Hedda, and that this allowed her to continue speaking Norwegian to him even though he only addressed her in English. I, on the other hand, attempted to establish a more monolingual English context during our recording sessions. Initially this was less aggressive, with a combination of the expressed guess strategy and the minimal grasp strategy. My use of these strategies during the first several sessions was not necessarily a conscious decision, however, and I have already discussed in chapter 3 that I used the expressed guess strategy much more often than I realized until the transcription stage, and in ways in which the child would be able to interpret that I had understood her Norwegian utterance. This is the challenge that meets the bilingual speaker, as establishing a truly monolingual context, especially in the home, can prove difficult. As the recording period progressed, however, I became more deliberate in my use of discourse strategies and began to employ the minimal grasp strategy more often. Typical

responses employing this strategy included “what?”, “I don’t understand,” and “I don’t know that word,” all of which signal to the child to varying degrees that her Norwegian utterance has not been met with comprehension. Additionally, explicitly asking for the utterance or word to be repeated in English was introduced as part of this strategy, although this response may indicate that the Norwegian was understood, but not the response I was looking for. Hedda’s mother also began to use this strategy when speaking English to Hedda during our sessions. Examples can be seen in (15). This strategy did not always elicit a response in English, however, and there was a marked difference between Session 14 and the earlier sessions in this regard.

(15a) (Session 4, 2;6.10)
*INV: what’s that?
*CHI: [- nor] &ha hund.
dog
‘Dog.’
*INV: what is it?
*CHI: dog.
*INV: yeah it’s a dog.

(15b)
*INV: what’s that?
*CHI: [- nor] hest.
horse
‘Horse.’
*INV: what is it?
*CHI: [- nor] hest.
horse
‘Horse.’
*INV: is it a horse?

(15a) shows an early example of a successful use of the minimal grasp strategy to elicit an English response, while (15b) shows a similar example from the same session where the minimal grasp strategy was insufficient and the investigator switched to the expressed guess strategy to keep the conversation going.

By Session 9, Hedda appears to be mimicking the discourse strategy employed by her mother and the investigator in asking if the investigator knows the Norwegian names of things - though whether she is consciously trying to establish a monolingual Norwegian context is uncertain. This session contains repeated instances of Hedda asking *Vet du ka den her hete?* (or ‘Do you know what this here is called?’ in English). This happens most often when Hedda

and the investigator are engaged in looking at picture books and the investigator asks Hedda to describe what she sees. An example follows.

(16) (Session 9, 2;11.17)
*INV: do you know what &um (..) do you know what this is?
*CHI: [- nor] ei ugle.
a owl
'An owl.'
*INV: do you know what that's called in English?
*CHI: a owl.
*INV: yeah it's an owl.
*CHI: [- nor] vet du ka det her hete?
know you what this here is-called
'Do you know what this here is called?'
*CHI: [- nor] det hete reinsdyr og Anna og Elsa.
it is-called reindeer and Anna and Elsa.
'It's called reindeer and Anna and Elsa.'

By Session 11 Hedda responds with *æ vet ikke* ('I don't know') for the first time when asked what the English word is for something. She shows an increased sensitivity to the language of her interlocutor and for the first time there are clear cases where she recognizes that she should respond with the English word and she doesn't know it, as in:

(17) (Session 11, 3;1.5)
*INV: yeah <what color> [/] what color is (..) this?
*CHI: [- nor] æ vet ikke.
I know not
'I don't know.'
*MOT: which [//] what's it called in Norwegian?
*CHI: [- nor] rosa.
pink
'Pink.'
*MOT: yeah and what's it in English?

In this case she eventually had to be told the English word, since she did not remember it. But what makes this example stand out from previous recordings is that instead of responding with the Norwegian word for the color (which she clearly knew) when unable to retrieve the English word in her lexicon, she expressed that she did not know the word in English.

Session 14 again stands out from the other sessions in that discourse strategies being used to try and invoke a monolingual English mode worked better and more quickly across the board. The discourse strategies themselves weren't being used in a different way, however, and there

are two factors that likely influenced their effectiveness instead. The first is of course the week of much more intensive regular input in English than Hedda was used to during her grandmother's visit. The second factor was that much of this session was conducted in a new area of the house compared with previous sessions – Hedda's brand new bedroom. This room had been painted a week or so before and was being prepared for Hedda to move into it, so it was new to her as well as to the investigator. This was the first session in which Hedda and the investigator engaged in play away from the main common areas of the house where guests are usually received – playing in Hedda's new room meant we were playing on a different floor of the house in a room with a door that closed.

Some examples follow, and it's apparent that the minimal grasp strategy in particular has an immediate effect not always seen in earlier recording sessions.

(18) (Session 14, 3;3.2)
*CHI: [- nor] du &k kan sove her nede.
you can sleep here down
'You can sleep down here.'
*INV: what?
*CHI: you can sleep under **mæ**.
me

(19)
*CHI: [- nor] vi må gå ut og se.
we must go out and see
'We have to go out and see.'
*INV: what are we doing?
*CHI: we gonna (.) go out (..) and see.

(20)
*INV: are we safe?
*CHI: [- nor] ja (.) vi skal lukke den her døra også.
yes we will close that here door-the also
'Yes, we'll close that door here too.'
*INV: what [/] what are we doing?
*CHI: we close the door.
*INV: we're closing the door too okay.

(21)
*CHI: [- nor] nu er det noen som banke på døra.
now is it someone who knocks on door-the
'Now there's someone knocking on the door.'
*INV: what?
*INV: what's happening?
*CHI: that's somebody **banking** on my +...

*INV: someone knocking on the door?
*CHI: yeah.

Additionally, in some cases Hedda catches herself before completing the Norwegian utterance, and switches to English without being prompted by any response from the investigator at all, even when she clearly wasn't comfortable with using English to express the thought.

(22)
*CHI: [- nor] du må sov +…
you must sleep
'You have to sleep…'
*CHI: you must be (...) in this (.) down (.) bunk.

(23)
*CHI: [- nor] <det var> [/] det var natt her ute.
it was it was night here out
'It was night out here.'
*CHI: it's (.) night.

Later transcripts will have to be examined in order to determine if this effect is sustained over time, even after Hedda's balance of input returned to normal (although subsequent recording sessions yet to be transcribed as well as discussions with her parents indicate that there may be a sustained effect). Nonetheless at this point the implications of her different behavior with regard to language choice and discourse strategy seems to point at the importance of input, and particularly the quantity of input.

Another study that potentially indicates a link between the effectiveness of discourse strategies and the importance of input is Slavkov (2015), previously discussed in Chapter 2. The child in Slavkov's study, Sophie, responded very well to discourse strategies used by her father to invoke a monolingual Bulgarian context (the minority language in Sophie's case). When Sophie's input situation changed dramatically with the introduction of half-day English-speaking daycare, however, she began to produce less Bulgarian and did not always respond to her father's attempts to establish a monolingual Bulgarian mode through discourse strategies. She was effectively becoming a passive bilingual as her Bulgarian underwent attrition. Ultimately, this led her father to adjust his own discourse strategies, using the repetition strategy or the move on strategy more than the minimal grasp strategy or explicitly asking for the word or utterance in Bulgarian. As discussed in Chapter 2, Sophie's production

did change again after a trip with her father to Bulgaria, indicating the importance of input. As Slavkov states, "[t]he trip to Bulgaria can be viewed not only as a powerful social factor in Sophie's attrition reversal but also as an important opportunity for exposure to richer input that she was able to take in over a short period of time, and derive potentially long-lasting benefits from it." It will be interesting to see in future work whether intensive periods of increased English input have a sustained or long-term effect on Hedda's language production as well.

Lanza (1992, 1997) points at a different relationship – that between the use of discourse strategies and intra-sentential code mixing. As discussed in Section 2.4, Lanza turned to analysis of discourse strategies when she found that the subject of her study, Siri, code mixed more when speaking with her father, who spoke her dominant language, than with her mother, who spoke her weaker one. Based on Hedda's case study this might seem surprising, and indeed code mixing more in the dominant language would be unexpected at first glance. Lanza examined the discourse strategies of both parents in the hope of finding an explanation for this, however, and she found that Siri's mother employed discourse strategies that signaled a monolingual English mode, while Siri's father tended to employ strategies that encouraged a bilingual context and strengthened his own bilingual identity. While Siri's parents for the most part followed the 1P/1L strategy, Lanza notes that Siri's father "actively used the repetition strategy, but also frequently modelled Siri's English words even though he would also supply the Norwegian equivalents," which reinforced his bilingual identity to Siri and the acceptability of a bilingual context. Hedda's father also recasts her Norwegian words sometimes, as seen in the following examples:

(24) (Session 1, 2;3.10)

*FAT: what's this?
*CHI: [- nor] det er gaffel!
*FAT: it's a **gaffel**, it's a fork.
*INV: a fork!
*FAT: and what's that?
*CHI: [- nor] det er kniv!
*FAT: a **kniv** is a knife.
*FAT: and what's that?
*CHI: [- nor] det er (.) skje!
*FAT: a **skje** is a spoon!
*FAT: that's right.

(25) (Session 8, 2;10.20)

*CHI: [- nor] æ plaske i dag (.) **in daycare**.
I splash today
'I splashed today…in daycare.'
*FAT: what did you splash?
*CHI: [- nor] nei.
*FAT: did you say **plaske**?
*CHI: [- nor] ja.

The first example is indicative of a frequent occurrence when the investigator was present, however, as Hedda's father often translated Hedda's Norwegian utterances for the investigator, as in the following example. The English sessions recorded with Hedda and her father when the investigator was not present would need to be examined in order to determine whether Hedda's father normally does this as often as he appears to in the sessions used for this thesis.

(26) Hedda and the investigator are looking at a set of picture cards together. Hedda's parents are nearby in the next room, within earshot.

(Session 3, 2;4.21)

*INV: what do you think they're drinking?
*CHI: [- nor] dem **drinking** &eh (.) sugerør.
they straw
'They're drinking…straw.'
*INV: okay.
*CHI: [- nor] sugerør.
straw
'Straw.'
*CHI: [- nor] han xxx.
he xxx
*FAT: **sugerør** is a straw.
*INV: oh.

Clearly, Hedda's code mixing, her input situation, and the discourse strategies employed by her interlocutors are all interrelated. Not only is the amount of input a factor on her code mixing in English versus in Norwegian, but the amount of input also effects how she responds to discourse strategies meant to establish a monolingual context in her weaker language, English.

4.4 Session 14 and cross-linguistic influence

Because the last period of increased English input seems to have had the largest sustained effect on Hedda's language development, I wanted to draw attention to and examine some of

the phenomena present in this transcript more closely. This file is particularly interesting because there are many things happening that aren't as prevalent (or are absent completely) in the previous transcripts. This is one of two sessions where the number of entirely English utterances is greater than the number of Norwegian utterances (as discussed in Section 4.1.3), and it is the only session in which Hedda used a greater variety of English words than Norwegian words (as discussed in Section 4.1.1). Additionally, many of Hedda's mixed utterances during this session were the result of instances where she said something in Norwegian first, the investigator responded with "what?", and Hedda tried her best to reframe the utterance in English, even if she was missing some lexical items. There are several examples of this, as in the following:

(27a)
*CHI: [- nor] nu er det natt.
now is it night
'Now it's night.'
*INV: what?
*CHI: <**nu** is it> [/] (...) <**nu** is it night.
now is it now
'Now it's night.'

(27b)
*CHI: [- nor] nu (.) er det morn!
now is it morning
'Now it's morning!'
*INV: what?
*CHI: [- nor] nu er det **Sunday**.
now is it
'Now it's Sunday.'

In the first example, Hedda is able to produce all but one word of her original utterance in English when she repeats it, although she maintains Norwegian word order. In the second example, she is less successful, and apparently unable to recall or produce the English word *morning*, she replaces it entirely with a different English word (the recording was conducted on a weekday, not a Sunday).

The following sections will examine a number of other phenomena present in this session and discuss the possibility of cross-linguistic influence in Hedda's English. As discussed in Chapter 2, Paradis and Genesee (1996) describe CLI/interdependence "as being the systemic influence of the grammar of one language on the grammar of the other language during

acquisition" at the level of competence. By Müller's (1998) definition, an ambiguous structure in one language overlaps with an unambiguous structure in the other language, and transfer occurs when the child makes use of the unambiguous structure from one language to deal with the ambiguity in the recipient language. This may not be the ideal way to describe the difference between verb placement in Norwegian and English described in Section 2.6.1, but as we will see in the next section there is a transfer effect.

4.4.1 Verb movement

Another thing that becomes much clearer in the examples from this session is the influence of Norwegian syntax on Hedda's production in English. In general, Hedda made much more of an effort to choose to use English, even when she appeared to feel uncertain of the exact target form she should be using. This can be seen in the previous two examples (27a-b), where Hedda's attempt to recast the utterance in English maintains V2 word order. Both of those examples were mixed utterances, but Norwegian syntax is present even in utterances where all of the words are actually English.

(28)
*INV: are you gonna look and see if there's a bear outside?
*CHI: [- nor] ja.
yes
'Yes.'
*INV: <okay I'll be quiet> [=! whispers].
*INV: <what do you see> [=! whispers]?
*CHI: I see not a bear.

(29)
*INV: do you have a pillow?
*CHI: [- nor] ka?
what
'What?'
*INV: do you have a pillow?
*CHI: yes!
*INV: yes.
*CHI: have you a blanket?
*INV: a blanket (.) do I need a blanket?

In the first example, Hedda employs the word *not* to negate her sentence, placing it after the verb where it would be in the Norwegian target sentence in her dialect (*æ ser ikke en bjørn,* or literally 'I see not a bear'). The target form in English would be *I don't see a bear*, making

use of *do*-support, and this type of error is almost never made by monolingual children acquiring English (Déprez and Pierce 1993). In the second example, Hedda asks the question *have you a blanket?* which also lacks *do*-support despite the fact that this has just been modeled by the investigator's previous question, asked twice in a row. The word order she uses again matches the word order for the Norwegian target by making use of inversion of the main verb and the subject, *har du et teppe?* (literally 'have you a blanket'). In fact, both of these word orders can be attributed to the V2 properties of Norwegian, like the examples in (27a-b) in the previous section.

Given that all verbs are subject to V2 in Norwegian in all three contexts shown in (27a-b), (28), and (29), and in English an auxiliary is sometimes needed and sometimes not to fulfill a V2-like pattern, Norwegian verb movement can be considered more systematic in that it is less ambiguous than the English pattern. This would fulfill Müller's (1998) criteria for cross-linguistic influence as well. Given that Hedda seems not to have acquired *do*-support which would be necessary for the target English forms in (28)-(29), this type of transfer may even be predictable. Bentzen (2000) also found evidence of transfer of Norwegian V2 properties into English in her study of a bilingual child acquiring English and Norwegian, and Westergaard's (2003) study of word order in young L2 learners of English whose L1 was Norwegian found a striking degree of transfer of V2 into English.

4.4.2 "Must" and other possible lexical interference

Hedda also repeatedly uses the word *must* where a native English speaker would be more likely to use a phrase like *have to* or *should*, and in some cases she uses the puzzling *must be* instead. This can be seen in the following examples, some of which have been discussed in this thesis already. In the last example, it's possible to see that she does also use other more target-like phrases such as *have to* in addition to *must*.

(30)
(a)
*CHI: I must **slå** off the lightings balls.
turn off
'I have to turn off the lighting balls [string lights with round paper shades].'

(b)
*CHI: you must be (…) in this (.) down (.) bunk.

(c)
*CHI: you must be **sov** under **meg**.
sleep me
'You have to sleep under me.'

(d)
*CHI: **du** must (.) open the door.
you
'You should open the door.'

(e)
*CHI: <I must be> [//] (.) I [/] I have to have the key!

'Must' is the translation equivalent for the very frequent Norwegian auxiliary *må*, which Hedda also uses in this session:

(31)
(a)
*CHI: [- nor] æ må våkne.
I must wake
'I have to wake up.'

(b)
*CHI: [- nor] vi må gå ut og se.
we must go out and see
'We have to go out and see.

(c)
*CHI: [- nor] nei æ må si det.
no I must say it
'No I have to say it.'

I propose that the frequency of *må* in Norwegian has influenced Hedda's lexical choice with regard to her frequent use of "must."

Another instance where Norwegian influence seems to have interfered with Hedda's lexical choice in English involves the case of a very curious word, rendered in its phonetic form as [ʃərɛnʧ], and rendered in the transcription as *&shi-wrench*. The main exchange where this word appears follows, with Hedda and the investigator discussing the troll they are pretending is outside Hedda's door:

(32)
*INV: okay what's that?
*CHI: [- nor] den man kan &shi-wrench-e trollet i fengslet.
that one can troll-the in jail-the
*INV: what [/] (.) what can you do with it?
*CHI: [- nor] &shi-wrench (.) æ knute han &eh (.) æ har en &shi-wrench.
I knot him I have a
*CHI: I have a &shi-wrench.
*INV: a [/] a wrench?
*CHI: yes.
*INV: you have a wrench what are you gonna do with it?
*CHI: no (.) <I'm gonna> [//] I have a key!
*INV: a key?
*INV: oh what are you going to do with the key?
*CHI: I'm gonna lock the door with it.

The investigator used the expressed guess strategy to try and make sense of the nonce form, which sounded like the English word *wrench* with the addition of the initial voiceless palato-alveolar sibilant. The actual item Hedda was referring to was a spare drawer pull, which both Hedda and the investigator pretended was a key they used to "lock" the door during the session. By the end of the exchange, Hedda has finally accessed and produced the word *key*. One plausible explanation for this form has to do with the Norwegian words for *key* and *wrench – nøkkel* and *skiftnøkkel*, respectively (the latter having the onset [ʃ]).

Determining whether or not these examples (Hedda's use of *must* as well as the bizarre wrench/key/*skiftnøkkel* blend) constitute cross-linguistic influence at the level of competence or if they are phenomena occurring at the processing level is beyond the scope of this thesis, but they are brought up here because they appear in conjunction with a burst of English production after a weeklong period of intensive English input. These should be kept in mind in the future as the subsequent sessions are transcribed and analyzed in case they (or similar phenomena) are observed again.

4.5 Conclusion

This chapter has presented an overview of the data compiled so far based on the transcripts in Hedda's corpus, as well as discussion of the implications of the data. Hedda's dominance in Norwegian was established on the basis of several measures, including the unique number of words per language used in each session, her mean length of utterance, and the proportion of

utterances in each language. This was used to successfully predict the directionality of her mixing following Lanza's (1992, 1997) observations about dominance and language mixing. Hedda's code mixing was broken down into categories to see how her mixing patterns compared from category to category. The effect of parental discourse strategies was also discussed, and evidence for cross-linguistic influence and language transfer in Hedda's English was outlined. Though the results should be considered preliminary due to the ongoing nature of the corpus project and the methodological concerns raised in Section 3.3, they corroborate the findings of many other studies of BFLA, particularly studies focused on typologically similar pairs (such as English + a Germanic V2 language). The preliminary findings will build the foundation for future work on the Hedda corpus.

5 Conclusion

The focus of this thesis has been code mixing by a bilingual child, acquiring Norwegian and English simultaneously. New (and ongoing) corpus data was presented and an analysis of code mixing data was carried out. The aim of studying this child's code mixing was to investigate which factors influence her mixing habits. While the claim of a unitary system in the early stages of acquisition has largely been rejected by the BFLA research community, Hedda's data provides continued evidence against it, as her mixing patterns show that she differentiates between different language contexts and uses her languages differently within them. Hedda's strong dominance in one of her languages also makes her an interesting case study – while it can make certain phenomena difficult to study because she produces less English than children like Siri (Lanza 1992) or Kate (De Houwer 1990), she does represent a point on the spectrum of dominance where many bilingual children fall, and learning more about the factors that affect strong dominance has implications for bilingual research. Hedda may be more sensitive to shifts in her input situation than a child like Siri, for example, which would mean bilingual children exhibiting strong dominance would make ideal subjects for studying subtle shifts in input. The introduction of Hedda's dense corpus constitutes a contribution of considerable size to the field of language acquisition research.

The factors I was most interested in examining with regard to Hedda's code mixing were her language dominance, discourse strategies, and language modes, all of which turn out to be interrelated. One of the biggest driving factors in Hedda's strong Norwegian dominance is input, a topic the BFLA research community has been showing a growing interest in examining in greater detail for several years (Paradis & Grüter 2014). Discourse strategies combine with input to amplify the effects of dominance, and they also help to establish monolingual or bilingual language modes. In Hedda's case, the imbalance between input in Norwegian and English intersects with discourse strategies, as her sole regular source of input in English, her father, does not make use of discourse strategies that are most effective at establishing a monolingual language mode. All of this combines to yield a child with strong Norwegian dominance. The interconnectedness of all of these factors points back to Grosjean's (2008) emphasis on a holistic view of bilingualism, in which the complex nature of being bilingual is recognized and the notion of the bilingual as "two monolinguals in one" is soundly rejected.

As this thesis project draws to a close, I am encouraged by the knowledge that the project of building Hedda's corpus is still only in its early stages, and that research on Hedda's language production will continue. It is my hope that the preliminary data and analysis contained in these pages will be able to aid other researchers interested in working with the corpus in the future, and I look forward to more detailed analysis of the topics covered here as more data becomes available. While one of the shortcomings of case studies is that they cannot be representative of a larger population on their own, the trade off is that all aspects of a case study can be examined in more detail than larger-scale studies allow. As interest in input as a research topic continues to grow, both large-scale studies such as De Houwer's (2007) survey as well as case studies such as this one will be increasingly important, for each can tell us something different about how language is acquired. It's the kind of knowledge that everyone benefits from, as new findings in acquisition research have implications not only for the field of linguistics, but also for fields like psychology, education, and speech and language pathology.

Works cited

Anderssen, Merete, & Bentzen, Kristine (2013). Cross-linguistic influence outside the syntax-pragmatics interface: a case study of the acquisition of definiteness. *Studia Linguistica*, 67(1), 82-100.

Bellugi, Ursula. 1967. *The acquisition of the system of negation in children's speech*. Ph.D. dissertation, Harvard University.

Bentzen, Kristine (2000). *I like not it like du like it! A case study of language transfer in bilingual first language acquisition.* Masters thesis, University of Tromsø.

De Houwer, Annick. (1990). *The acquisition of two languages from birth: A case study*. Cambridge University Press.

De Houwer, Annick. (2003). Home languages spoken in officially monolingual Flanders: A survey. *Plurilingua*, 24, 71-87.

De Houwer, Annick (2007). Parental language input patterns and children's bilingual use. *Applied psycholinguistics*, 28(03), 411-424.

De Houwer, Annick. (2009). *Bilingual first language acquisition*. Multilingual Matters.

De Houwer, Annick (2014). The absolute frequency of maternal input to bilingual and monolingual children. *Input and experience in bilingual development*, 13, 37.

Déprez, Viviane, & Pierce, Amy (1993). Negation and functional projections in early grammar. Linguistic Inquiry, 25-67.

Genesee, Fred (1989). Early bilingual development: one language or two? *Journal of child language*, 16(01), 161-179.

Genesee, Fred (2008). Bilingual first language acquisition: Evidence from Montreal. *Diversité urbaine*, 9-26.

Genesee, Fred (2016). 1 Shifting Perspectives on Bilingualism. *Bilingualism Across the Lifespan: Factors Moderating Language Proficiency*.

Greene, Kai J., Peña, Elizabeth D., & Bedore, Lisa M. (2013). Lexical choice and language selection in bilingual preschoolers. *Child Language Teaching and Therapy*, 29(1), 27-39.

Grosjean, François (1982*). Life with two languages: An introduction to bilingualism.* Harvard University Press.

Grosjean, François (1998). Studying bilinguals: methodological and conceptual issues. *Bilingualism: Language and cognition*, 1(02), 131-149.

Grosjean, François (2008). *Studying bilinguals*. Oxford University Press.

Grüter, Theres, et al. (2014). Language exposure and online processing efficiency in bilingual development. *Input and experience in bilingual development*, 15-36.

Hulk, Aafke & Müller, Natascha. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and cognition,* 3.3:227–244.

Lanza, Elizabeth (1992). Can bilingual two-year-olds code-switch? *Journal of child language*, 19(03), 633-658.

Lanza, Elizabeth (1997). *Language Mixing in Infant Bilingualism: A Sociolinguistic Perspective*. Oxford: Clarendon Press.

MacWhinney, Brian (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates

Meisel, J. M. (1989). Early differentiation of languages in bilingual children. *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss*, 13-40.

Müller, Natascha (1998). Transfer in bilingual first language acquisition. *Bilingualism: Language and cognition*, 1(03), 151-171.

Müller, Natascha, & Hulk, Aafke (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and cognition*, 4(01), 1-21.

Paradis, Johanne, & Genesee, Fred (1996). Syntactic acquisition in bilingual children. *Studies in second language acquisition*, 18(01), 1-25.

Paradis, Johanne, & Grüter, Theres. (2014). Introduction to 'Input and experience in bilingual development.' *Input and experience in bilingual development*, 1-14.

Paradis, Johanne, Crago, Martha, Genesee, Fred, & Rice, Mabel (2003). French-English Bilingual Children With SLI: How Do They Compare With Their Monolingual Peers? *Journal of Speech, Language, and Hearing Research*, 46(1), 113-127.

Platzack, Christer. (1996). The initial hypothesis of syntax. *Generative perspectives on language acquisition*, 369-414.

Rodina, Yulia, & Westergaard, Marit (2015). Grammatical gender in Norwegian: Language acquisition and language change. *Journal of Germanic Linguistics*, 27(02), 145-187.

Santelmann, Lynn et al. (2002). Continuity and development in the acquisition of inversion in yes/no questions: dissociating movement and inflection. *Journal of Child Language*, 29(04), 813-842.

Santelmann, Lynn (1995). *The acquisition of verb second grammar in child Swedish: Continuity of universal grammar in Wh-questions, topicalization and verb raising* (Doctoral dissertation, Ohio State University).

Slavkov, Nikolay (2015). Language attrition and reactivation in the context of bilingual first language acquisition. *International Journal of Bilingual Education and Bilingualism*, 18(6), 715-734.]

Swain, Merrill. (1972). *Bilingualism as a first language* (Doctoral dissertation, University of California).

Thornton, Rosalind, & Tesan, Graciela. (2013). Sentential negation in early child English. *Journal of Linguistics*, 49(02), 367-411.

Unsworth, Sharon et al. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35:04, 765-805.

Volterra, Virginia and Taeschner, Traute (1978). The acquisition and development of language by bilingual children. *Journal of Child Language*, 5(02), 311-326.

Westergaard, Marit (2003). Unlearning V2: Transfer, markedness, and the importance of input cues in the acquisition of word order in English by Norwegian children. *EUROSLA yearbook*, 3(1), 77-101.

Westergaard, Marit. (2005). *The development of word order in Norwegian child language: the interaction of input and economy principles in the acquisition of V2*. University of Tromsø.

Yow, W. Quin, Patrycia, Ferninda, & Flynn, Suzanne. (2016). Code-switching in Childhood. *Input and Experience in Bilingual Development*.

Zeijlstra, Hedde. (2004). Sentential negation and negative concord. LOT/ACLC.

Zeijlstra, Hedde. (2008). Negative concord is syntactic agreement. Ms., University of Amsterdam, 5, 113.