A Corpus-Based Study of Discourse Markers in English as a Second Language by Arabic-Speaking Children

Nuha Alharbi

A thesis submitted for the degree of Doctor of Philosophy in Linguistics

Department of Linguistics and English Language

May 2025

Declaration

I declare that this thesis is the result of my own work and has not been submitted in support of any application for another degree at this or any other university. I confirm that all materials have been properly acknowledged where reference has been made to the work of others.

Abstract

This thesis presents a corpus-based approach to the exploration of the use of discourse markers in the speech of Arabic-speaking school-age children learning English as a second language. The study compares the use of discourse markers in this corpus of Arab children with a corpus of the language of English-speaking children. In addition, it examines the role of context and length of preschool study on the use of English discourse markers by Arab children.

Using both quantitative and qualitative approaches, the present study surveys the frequencies and functions of English discourse markers in the speech of Arabic-speaking children in comparison to their usage in age-matched English-speaking children. In order to address the multifunctionality of discourse markers, the study adopts a proposed framework of functions based on previous studies (Müller, 2005; Aijmer, 2013; Brinton, 1996, 2006, 2017, 2023). The framework consists of two domains of functions, i.e. textual and interpersonal, in which a set of subfunctions is listed.

The eight most frequent discourse markers in the two corpora are identified. Quantitative analysis shows that the discourse markers *like*, *yeah* and *okay* are the most frequent discourse markers in the speech of Arab children. English-speaking children use *yeah*, *like* and *oh* more than the other discourse markers. Both groups of children rarely use the two-word discourse markers *you know* and *I mean*. Among the eight selected discourse markers studied, *well* is comparatively the least frequent in the speech of Arabic-speaking children. Analysing the frequencies of discourse markers in the speech of Arabic-speaking children according to the context and duration of learning English as a second language shows no differences in the

overall use of discourse markers. However, there is some variation in frequency of some discourse markers between the groups of Arab speakers.

Subsequent qualitative analysis of the eight discourse markers offers some indications of differences between the use of discourse markers by Arabic and English-speaking children. These differences appear according to the frequencies of discourse marker functions between the two groups of speakers. In addition, each group of children uses some discourse markers to fulfil certain functions that are not performed by the other group.

Acknowledgments

This thesis would not have been possible without much help from many people whom I must acknowledge. First and foremost, I must thank my two supervisors, Dr Luke Collins and Dr Andrew Hardie. Dr Collins's continuous guidance and enormous assistance have proved invaluable throughout my study. My words cannot express how I really appreciate all that he has done to guide me through this long journey. I would also like to extend my thanks to Dr Hardie for his assistance and guidance in writing my thesis. Dr Hardie has inspired me a lot with his passion to work with corpus linguistics and with his willingness to offer his guidance when needed. Indeed, I appreciate his assistance in constructing my own corpus and uploading it to the CQPweb server. Dr Collins and Dr Hardie are truly mentors and role models who have made this thesis possible and to whom I owe a world of thanks.

Many people have contributed to the completion of my doctoral studies. Dr Dana Gablasova offered advice related to data collection, in particular, during my pre-confirmation panel. Thanks to her, I managed to organize my data collection, taking into account the amount of data and time required to achieve my research aims. Her advice saved me time and effort during my data collection. I am also thankful to Dr Isobelle Clarke for her insightful comments during the confirmation panel. I must also extend my thanks to the school leaders for their assistance in facilitating the study and to the children's families for consenting to the data collection process and offering help where they could. Critically, this thesis would not have been completed without the generous contributions of the schoolgirls who willingly shared their thoughts with their friends in the conversations that make up the data analyzed in this study.

I would like to thank all the people who have been there for me during the years of my study. First, I would like to thank the man who always stands with me, keeping his promises to be with me through my ups and downs: my husband, Thabet. To my children, Alin, Yasameen, Ruba and Yazeed, thanks for being in my life, for inspiring me and for sharing this long journey with me. I would also like to thank my parents, my sisters and my brothers for their unfailing support, for their prayers and for their understanding of my absence from their gatherings. Last but not least, I owe a special thanks to my sister, Samar, for sacrificing her time and being there for me during my hardest days. This list could be much longer, and I will be always thankful for everyone who has offered me any kind of help, even in the smallest way.

Contents

List of Tables

List of Figures

List of Abbreviations

ESL	English as a second language
CL	Corpus linguistics
DM	Discourse marker
L1	First language
L2	Second language
PM	Pragmatic marker
SLA	Second language acquisition

Chapter 1: Introduction

1.1. Motivation of the thesis

This thesis presents a study in which I employ corpus methods to investigate Arabic-speaking children's pragmatic competence in English as a second language (ESL) through their use of English discourse markers (DMs), and to compare their usage with that of English children. Examining language competency goes beyond studying acquisition of grammatical constructions. It is also important to consider learners' appropriate use of language in social interactions, that is, *pragmatic competence* (Taguchi, 2019: 1). Pragmatic competence in a second language (L2) entails being able to modify one's speech and to reconstruct a communicative act according to a listener's reaction (Culpeper et al., 2018: 6). Pragmatic competence also comprises the ability to master connectivity in ongoing communication: what is referred to as 'pragmatic fluency' (House, 1996: 228). Under this term, House (1996) lists certain language behaviours – adopted from Riggenbach (1991) – as fluency-related phenomena in L2 learners. They include the ability to carry on a discourse; the ability to initiate and change topics; and the use of pauses and repairs when necessary (Riggenbach, 1991: 439). These language behaviours are most explicitly manifest through the use of DMs.

Many studies show that children are able to use DMs in their first language (L1) as indicators of pragmatic competence from an early age, i.e. three years old (see Kyratzis & Ervin-Tripp, 1999; Escalera, 2002, 2009). In the field of L2 acquisition (SLA), only a few studies have focused on L2 children's use of DMs, although there are a plethora of studies examining the acquisition of DMs in adult L2 learners' speech (Jakupčević, 2019: 412), for example, the speech of university students. But it seems very probable that investigating these markers will provide an indication of children's level of development in English as L2

speakers, as a marker of their pragmatic competence. The present study aims to address the consequent gap in the literature by investigating the use of DMs in the speech of Arab children learning English as a L2.

DMs are small words that link different parts of discourse and assist speakers in their interaction with their listeners (Van Olmen & Šinkuniene, 2021:1). They represent a significant class of elements in linguistic communication, described not only as the 'glue' that binds a preceding utterance to a following one (Fraser, 2021: 314), but also as discourse organizing devices or as transitional or boundary markers in speech (Andersen, 2011: 599; Fraser, 2021: 315). They make "a limited contribution to propositional content, but contribute considerably at pragmatic level" (Andersen, 2011: 599). Mosegaard Hansen (1998: 199) argues that misuse of DMs by L2 speakers will result in misunderstandings in communication, in ways that are different from misunderstanding caused by incorrect use of content words.

DMs are often overlooked by L2 instructors. Teachers are mostly concerned with achieving lexical/grammatical proficiency in their students' speech. This focus may lead them to ignore the importance of enhancing students' pragmatic competency, that is, their facility in using features such as DMs. Crible and Pascual (2020: 56) assert that because of a scarcity of explicit teaching and limited natural communication settings, L2 learners may underuse or misuse DMs. Although the absence of a DM may not affect a sentence's construction, it can greatly affect the attitude communicated by the speaker towards others. A speaker who employs DMs is seen as "friendlier, politer, more educated, and more confident than the speaker omitting them" (Buysse & Blanchard, 2022: 240). L2 speakers often employ DMs to help clarify their implicit message and hence, decrease ambiguity in their speech (Fraser, 2021: 316).

Researchers agree that DMs are multifunctional elements in that they can serve different functions in different contexts. This multifunctionality adds to their significance in real-life interactions. However, this feature of DMs poses a challenge for L2 speakers to learn (Fischer, 2014: 271). While much published research discusses the use of English DMs by non-native adult speakers from various native language backgrounds (e.g. Müller, 2005; Fung & Carter, 2007; Asik & Cephe, 2013; Gilquin, 2016; and others), including Arabic (e.g. Kurdi, 2008; Al-Yaari et al., 2013; Al Makoshi, 2014; Algouzi, 2014), less is known about the use of DMs by child learners of English (Romero-Trillo, 2002; Jakupčević, 2019). Studies of Arabic-speaking children learning English as a L2 have tended to focus on phonological, morphological and grammatical development. There is a scarcity in investigating how DMs are being used by Arab children learning English as L2 in Arab countries.

In the Kingdom of Saudi Arabia, many Arab families enrol their children in English-medium schools due to the high quality of teaching in these schools (Alfaraidy, 2020: 229). These schools are seen as able to fulfil Arab parents' desire to equip their children with the education that they will require in addition to learning English as a world language (Al-Qahtani et al., 2016: 31). English-medium schools, also known as international schools in Saudi Arabia, presently offer courses at all educational levels (i.e. primary, intermediate and secondary). Until 2009, such schools were permitted to teach only children of English-speaking expatriates. But since then, the number of Arab children in these schools has increased dramatically, for instance rising from 162,615 in 2015 to 205,377 in 2021 (Ministry of Education, 2021). The enrolment of Arab children to international schools offers a rich site for English-language learning research. However, researchers who have to date investigated these schools have focused either on attitudes and perceptions towards using English as a medium of communication and teaching on the part of families and teachers (e.g. Al-Qahtani et al., 2016; Ismail & Shaban, 2016; Alfaraidy, 2020), or on the effect of English-medium

teaching on children's performance in other subjects such as science or maths (e.g. Jarrah, 2020). By contrast, there is a scarcity of research on how these schools support children in their English language skills. The present study seeks to close this gap by investigating how Arab children benefit from using English extensively in international schools, specifically in terms of their pragmatic competence and language use in communication with their peers in informal contexts. Furthermore, previous studies have shown that the contexts in which speakers learn English, such as their experiences studying in English-speaking countries, can influence the variation of DMs in their usage as adult speakers of English as a L2 (e.g. Polat, 2011; Gilquin, 2016; Ament & Parés, 2018; Magliacane & Howard, 2019; Diskin-Holdaway, 2021). As such, I will also consider the language histories of participants in my study, specifically their experience of English-language preschool – whether in Arabic-speaking countries or in English-speaking countries.

1.2. Objectives of the thesis

This study is designed to provide a comprehensive quantitative and qualitative analysis of the use of the most frequent DMs used by Arabic-speaking children learning ESL. Research on SLA has shown that corpus linguistics (CL) is an effective methodology for observing patterns of language use. Thus, it has been utilized in many prior studies of language acquisition and development (Paquot & Tracy-Ventura, 2020: 1). In this study, I create a corpus of interactions between Arabic-speaking children enrolled in English-medium schools. The compilation of a corpus of Arabic-speaking children participating in informal conversations in the present study allows us to look for elements of language use and observe their functions in semi-spontaneous interactions. To compare Arabic-speaking children's use of DMs with that of L1 speakers, I also create a customised sub-corpus of age-matched L1 English speakers taken from the Spoken BNC2014 (Love et al., 2017). This allows me to conduct a comparison of the uses and functions of frequent DMs in informal L2 peer-

conversations where English is the L1. Using corpus tools (e.g. wordlists and concordances), I analyse the functions of DMs through a top-down process, starting with a definitive framework for all DMs followed by a qualitative analysis of each DM's function in the corpus.

1.3. Hypotheses and research questions

The study aims to test hypotheses about frequencies and functions of English DMs in the speech of Arab children at English-medium international primary schools in Saudi Arabia. The two principal hypotheses in this study are:

(H1) The frequencies and functions of DMs in Arabic-speaking children's speech differ from that of native English-speaking children.

(H2) The frequency of DMs will be higher in the speech of children who have spent their preschool years in English-speaking countries, as opposed to English-medium preschool in Arab countries. Moreover, it is hypothesised that the longer the Arabic-speaking children have been exposed to English, the more varied their use of DMs will be.

The first hypothesis is motivated by previous studies that examine the use of DMs by L1 and L2 speakers. Many of these studies report that L2 speakers use DMs differently from L1 speakers either in that i) they use fewer types of DMs, ii) they use the same DMs with lower frequency, or iii) they misuse the functions of these markers. The small number of such studies that look at L2 children have also reported less frequent use of DMs. However, looking at DMs just in terms of overuse or underuse in L2 speech does not offer any indication of whether the speakers are making appropriate use of DMs. On the one hand, the overuse of DM frequency is not an indicator of successful learning of DMs since repeating the same DM in a conversation may be perceived by other participants as a hesitant signal, leading them to perceive that speaker as "unskilful" or "powerless" (Crible & Pasqual, 2020:

55). On the other hand, underusing DMs in the speech of learners does not prove their incompetency because L2 learners may use fewer DMs but show appropriate and creative use of their functions (House, 2013). Hence, it is important to look at the variety of functions fulfilled by these markers.

The second hypothesis is motivated by L2 researchers' view of the effect of external individual differences on the use of linguistic patterns and pragmatic features (see section 2.2). Indeed, many such external factors have been proven to have a great effect on learners' production. These factors include the context and length of time of the L2 speaker's study of the L2 (see section 2.3).

To test these hypotheses, three research questions will be addressed in the analysis:

1. Are there any differences between the frequencies of the most frequent DMs in the ESL speech of Arabic-speaking upper-grade children (fourth, fifth and sixth grades) in Saudi international primary schools, compared to age-matched native English-speaking children?
2. How does the context of formal exposure to English prior to joining primary school affect the frequency of DMs in the ESL speech of these Arabic-speaking children?
    a) Is there any difference in the frequencies of DMs between participants who have studied ESL in preschool in Arabic-speaking versus English-speaking countries?
    b) Is there any effect of the number of years of attendance at preschool in English-speaking countries on the frequencies of DMs?
3. What are the functions of the most frequent DMs in the speech of these Arabic-speaking school-age children? Are there any differences in the functions with which DMs are used, in comparison to native English-speaking children?

Answering these research questions will enhance our knowledge of how Arabic-speaking children acquire knowledge of DMs in English. The answers to these questions will also contribute to the literature on L2 pragmatic competence; since the existing studies focus mainly on adults' acquisition of these pragmatic features; an investigation of these features in the speech of children will fill a gap in what is presently known. Finally, it is hoped that the results regarding the effect of external factors on these children's use of DMs may contribute to future research on the field of child L2 pragmatic acquisition.

1.4. Outline of the thesis

The thesis is structured as follows. Chapter 2 reviews the literature on child SLA, with a focus on the effect of individual differences on pragmatic competence. It also discusses the long-standing theoretical debate regarding a unified definition for DMs, as well as their characteristics and functions, and reviews a number of studies of DMs undertaken within the field of SLA. The chapter then considers the importance of using corpus methodology in the field of SLA, introducing some well-known methods of corpus data analysis, including frequency lists and concordance lines.

Chapter 3 addresses the methodology of the study. It outlines the compilation of a corpus of English speech by Arabic-speaking children, covering the processes of data collection and transcription and coding of the data. The chapter then reports on the construction of a corpus of L1 child speech as a point of comparison by extracting text from an existing reference corpus, namely, the Spoken BNC2014. Finally, it discusses the steps of analysis applied to the data, including a justification for the statistical analyses utilised and a presentation of a framework of DM functions.

Chapter 4 presents an overall look at DM use in the two corpora. This begins with quantitative analysis, first of the frequencies of DMs generally, and then of the use of DMs in the speech of Arab children in relation to the context and duration of their study of English as

a L2. After the quantitative analysis, a qualitative analysis of each DM is undertaken. The DMs that are observed in the speech of Arab children with relatively low frequency are dealt with in this chapter, whereas the following two chapters (5 and 6) present a detailed qualitative analysis of the most frequent DMs in the two corpora.

In each of chapters 5 and 6, there is an introductory section which presents a review of literature of the DMs considered in that chapter. These reviews address the various functions identified in the literature for each DM. Then, detailed quantitative and qualitative analyses of the functions of each DM are presented, with examples from the two corpora of L1 and L2 children speech.

Chapter 7 is a discussion of the significant findings of chapters 4 to 6. The chapter also comments on the implications of the findings for directions of future studies on child SLA and the use of DMs. It also discusses some of the limitations of the present study and the implications of the present study on the field of child L2 and pragmatic competence.

1.5. Summary

This introductory chapter outlines the thesis. It begins with stating the motivation behind the present study and asserting the significance of studying the speech of Arabic-speaking children enrolling in English-medium international schools. Section 1.2 discusses the objectives of the present study. The main objective is to investigate features of pragmatic competency, namely, DMs in the speech of L2 children. Section 1.3 addresses the research questions and explains the hypotheses behind these questions. Section 1.4 highlights the structure of the present thesis, with some details about the contents of each chapter. The following chapter will present a survey of relevant literature on corpus linguistics, DMs and child SLA.

Chapter Two: Literature Review

2.1. Overview

In this chapter, I review literature relevant to the study of pragmatic language development among children, the study of DMs in children, and the use of corpus approaches to investigate discourse features in SLA. Thus, the chapter is divided into three main sections. In section 2.2, I explore the importance of investigating certain contextual factors in child SLA. Then, I discuss the significance of these factors in relation to children's acquisition of pragmatic competence. Section 2.3 presents perspectives from the literature on defining DMs and on their functions and distinctive features, before moving on to review a number of studies on DM acquisition in the field of SLA. Section 2.4 discusses the rationale for using corpus approaches in the present study. I also present some relevant literature on corpus methodology as used in the field of SLA in this section.

2.2. Child SLA

Child SLA is an emerging subfield of SLA, a field that studies the process of learning another language after acquiring the first language (also known as a native language, mother tongue or L1) by adults or children (Saville-Troike, 2012: 2). Recently, child SLA has come to be seen as a separate subfield of SLA (Philp et al., 2008: 4), distinct from research into simultaneous language acquisition, which is known as *bilingualism* (Paradis, 2007: 387). While *bilingualism* refers to the acquisition of two languages from birth, child SLA denotes the process of acquiring L2 in childhood after the establishment of an L1 (Paradis, 2007: 387). In other words, the time when children start to acquire the L2 determines the type of acquisition. Some researchers in the field of child SLA argue that the term *bilingual* is equivalent to L2 learner (Slabakova, 2016: 93). These researchers' argument is grounded on

Grosjean's (1989: 4) definition of bilingualism. This definition of bilingual speakers extends from minimum and medium proficient L2 speakers up to speakers who obtain full command of the two languages. Thus, bilingualism is seen "as a gradient measure, rather than a dichotomous category" (Serratrice, 2019: 16). Child SLA is thus seen as part of this gradient (Slabakova, 2016: 93). On the other hand, Haznedar & Gavruseva (2008: 3) argue that acquisition of an additional language from birth should be known as *simultaneous bilingualism*, while they characterise children learning a L2 after acquiring the basics of L1 as *consecutive bilinguals*. In this sense, we can refer to child SLA as a consecutive bilingual process in which a child learns the L2 after being exposed to L1 for some time. We can also agree that the time of first exposure to L2 can determine the type of L2 acquisition in children.

The contexts in which L2 children start to speak the L2 differ from that of bilingual children (Serratrice, 2019: 16). A L2 child often learns one language before the other, and moreover speaks the first language mainly at home and the additional language at school (Paradis, 2007: 387). Once a child or a toddler joins a preschool program, their language-learning environment begins to change. This change happens as they start to interact with people outside their homes. This change is one of many factors that researchers frequently consider while studying child SLA.

2.2.1. Factors affecting child SLA

In SLA, the factors that affect the acquisition process are known as individual differences (Dörnyei &Skehan, 2003; Ehrman et al., 2003; Dörnyei, 2005). These factors are also studied in the field of child SLA. Indeed, SLA in children is influenced by a myriad of external and internal factors (Paradis, 2011:213). The external factors include the socio-cultural environment, educational settings, and the length and amount of exposure to diverse

linguistic inputs. Internal factors include cognitive maturation, age, effect of L1 and motivation and each can play a crucial role in shaping a child's ability to master the L2 (Hoff, 2014: 276; see also Dörnyei, 2005: 2). Thus, investigating the relation between these external and internal factors and children's development of their pragmatic competence in a L2 can enhance our understanding of the nature of successful child L2 pragmatic acquisition.

2.2.1.1. Internal factors in child SLA

Internal factors are those that affect child SLA and are related to the child's personality, age, L1, motivation and aptitude. Age is a central individual factor and has been extensively studied in SLA, often from the perspective of the Critical Period Hypothesis. This hypothesis, as advanced by Lenneberg (1967), specifies that there is a certain period of life in which language can be acquired easily. It proposes that before puberty, learners face fewer obstacles in gaining native speaker competence (Ellis, 2015: 37), while achieving native-like *proficiency* in a language after puberty is much more difficult (Slabakova, 2016:85; see also Collier, 1987). Compared to adults, children can learn languages more easily, and become more proficient, due to the plasticity of their brains (Birdsong, 2018: 3). Studying the effect of age requires us to be able to compare adult and child proficiency. Therefore, a central concern in this area is how exactly proficiency level can be measured in a way that allows adults and children to be compared. Although *proficiency* is an essential measure in the study of child and adult SLA, it is also difficult to define. The definition and scope of proficiency vary according to the type of assessments and measurement applied (Ingram, 1998). A broad general definition of the term *proficiency* in SLA is that it refers to "a person's overall competence and ability to perform in L2" (Thomas, 1994: 330). Measuring proficiency in children and adults implies measuring competency from different linguistic aspects, including phonological, morphological, syntactic and discourse elements (Unsworth, 2008: 307). In comparing children's and adults' proficiency levels, researchers should match the right adult

group with the appropriate children. Highly proficient children should be compared with highly proficient adults and likewise for low proficiency language users (Haznedar & Garuseva, 2013: 352). Adult learners are seen to have an advantage over children in developing knowledge of vocabulary (Ellis, 2015: 43). In addition, adults are better at utilising conscious learning strategies or explicit learning. Comparing adult and children proficiency, Slabakova (2016:88) finds that adults may achieve native-like language in their daily communications, but still perform lower than the native speaker range in lexical, grammatical and perceptual abilities. Although adults are proven to be faster L2 learners, children can follow or surpass them in becoming more proficient L2 speakers (Ellis, 2015: 43). In addition, children can acquire better native-like accents in a L2 than adults (Collier, 1987:2; Ellis, 2015: 41). Hoff (2014: 50) argues that the earlier the language is acquired, the more native-like proficiency is gained.

The investigation of age-related competence in the L2 has broadened from looking at differences between adults and children to looking at differences among age groups of children. This is based on the idea that children vary in how they acquire their L1 and L2 according to the stages of cognitive development that they pass through (Philip et al., 2008: 5). Researchers typically use schooling levels as a way of separating children out into different stages of cognitive development. These levels include: a) early childhood (2-7 years), b) middle childhood (7-11 years), c) early adolescence (12-14 years), and d) later adolescence (15 years and older) (Philip et al., 2008: 5, see also De Houwer, 2021). Studies on child SLA often take these different stages into account in order to examine the validity of the notion that "earlier is better" in SLA (e.g. Blom & Bosma, 2016; Paradis et al., 2017). A younger age of acquisition does not always entail better or faster acquisition of a L2. Older children (in middle childhood) demonstrate positive results in SLA compared to their younger counterparts, due to cognitive and linguistic maturation (Paradis et al., 2017: 161;

see also Paradis, 2023). For instance, while children learn discourse strategies such as explanation from an early point (from three years according to Barieri, 1989), young children display little diversity in their explanation strategies. They rely mainly on repetition of information rather than justification or elaboration (Cook-Gumperz & Kyratzis, 2015: 599).

Children can learn conversational and discourse strategies regardless of their age in acquiring the L2. They can pick these patterns at any age. However, mastering complex patterns including politeness strategies can pose a challenge to young learners, since these patterns are often sensitive to age and social factors (Caldwell-Harries & MacWhinney, 2023: 11). As children grow older, they become more logical thinkers; linguistically speaking, they become more attuned to turn-taking management and use of pragmatic features including speech acts (Philip et al., 2008: 6). They also become competent in using pragmatic actions in L2, even if their syntactic skills and vocabulary are still limited (Alduais et al., 2022, see also Cekaite, 2007). Subsequently, middle age children may benefit from learning these complex strategies from their peers in a school context (Caldwell-Harries & MacWhinney, 2023: 8).

Another internal factor is the effect of L1. There exists a long-existing debate on how L1 affects the acquisition of other languages. The Contrastive Analysis approach (Lado, 1957 cited in Lennon, 2009: 51), which is rooted in the behaviourist theory[1], hypothesises that L2 learning is made difficult or easy by the nature of the L1 relative to the L2 (Lennon, 2009: 51). According to the behaviourist view, L2 acquisition is a process of establishing a new set of behaviours. The existing behaviour, that is the knowledge of L1, is the main factor that can hinder the acquisition of L2. Thus, unsuccessful SLA has been attributed to *transfer* from L1 (VanPatten & William, 2015: 20). The knowledge and properties of children's first language may affect the acquisition of a new one positively or a negatively. Ellis (2015: 129) explains

[1] The behaviourist view of language acquisition emerged from psychology theory to account for language acquisition as habit formation (Skinner, 1957).

positive and negative L1 transfer on SLA. In positive language transfer, the first language is seen as a *facilitator*, whereas, in negative transfer, the first language *interferes* with SLA (Ellis, 2015: 129). On one side, similarities between two languages may accelerate the acquisition of additional languages (Krashen, 1981: 69; see also Ellis, 2015, Ren, 2023). When a young child begins to learn two languages simultaneously, they will easily learn words that are similar in the two languages (De Houwer, 2021: 28). This is an example of positive transfer. On the other side, negative transfer occurs if the two languages are highly distinct. Language differences can lead to a higher rate of difficulties and errors in acquisition (Ellis, 2015: 134). Negative transfer between Arabic and English can occur due to many extensive differences between these two languages (Palmer et al., 2007: 12). For instance, in Arabic adjectives follow the nouns that they modify, which is the opposite order to that in English. In addition, unlike the SVO word order of English, Arabic utilises two different basic word orders: SVO and VSO. According to the contrastive approach, learners often transfer their knowledge of Arabic to English, producing ungrammatical English sentences. This generalisation of Arabic rules to English is an example of negative transfer (Albirini, 2018: 178).

Opponents of the behaviourist view of language learning as habit formation have criticised the view of the learner's first language as the main source of errors (Lightbown & Spada, 2006: 93). One aspect of such criticism is based on the view that children can produce utterances beyond the imitation process (VanPatten & William, 2015: 23), an idea presented by The Mentalist Theory advocates who emphasise looking at the mental capacities for language in children's brains. Another challenge to the behaviourist view is found in Chomsky (1965: 37), who proposes that children are equipped with a particular *Language Acquisition Device*. This device contains the universal linguistics knowledge of all languages from which they can construct the knowledge of their own language.

Another attempt to explain the effect of first language on SLA is the proposed hypothesis of the *interlanguage* system in the seminal work of Selinker (1972). According to Selinker (1972: 214), *interlanguage* is a linguistic system that contains all the utterances produced by a L2 learner in their attempts to produce native-like output. In explaining the interlanguage system, Selinker (1972: 215) introduces the concept of *fossilisation*, a concept that refers to a speaker's overgeneralisation of linguistic rules and items from L1 in their interlanguage system. Learners *fossilise* these linguistic items or rules in their communication in L2. Fossilisation may occur regardless of a learner's age or the input they have received. The development of any learner's interlanguage will stop at some point due to fossilisation, such that learners will never reach the level of competency associated with native speakers (Tarone, 2018: 1).

A contemporary functional approach to language acquisition that addresses the influence of L1 on SLA is the Usage-based approach. According to the usage-based theoretical framework, children often acquire linguistic elements such as DMs in social context, through repeated exposures from parents and others. While researchers adopting the earlier discussed approaches (i.e. behaviourist and mentalist) rely often on "constructed examples, introspective intuitions and judgment tasks", researchers adopting the usage-based theoretical approach use observations and corpus evidence to study child language acquisition (Zahler, 2023: 74). I will expand on the usage-based framework and its relation to the present study in section 2.2.2.

2.2.1.2. External factors in SLA

External factors that are relevant in child SLA include the social contexts in which the language is learned and the amount of input children receives in these social contexts. It is often typical for a child to acquire their L1 through parental interactions. In contrast, a L2

will be acquired in most cases by communicating with large communities inside and outside the home, such as with school teachers or peers (Unsworth, 2016b: 121). These communications, however, can take place in very different settings. In contrast to educational contexts, which are formal and conscious learning settings, naturalistic social contexts are informal settings in which learning is an unconscious process. Ellis (1994: 216) argues that naturalistic contexts yield higher proficiency and more grammatical competence in language than the educational context.

Educational settings include schools or preschools in which children have an opportunity to learn rules of communication in L1 as well as L2, such as how to speak formally to their teachers, or appropriate versus inappropriate times at which to interrupt teachers or classmates (Blum-Kulka & Snow, 2004: 328). Outside of lessons, children communicate with their peers during school time in more naturalistic settings (e.g. lunch time). These settings are also a locus of language learning. In particular, through peer interactions, children pass on linguistic knowledge acquired from adults in other contexts. Moreover, communicating with peers during school time is also how children learn to construct their own culture (Cook-Gumperz & Kyratzis, 2015: 599). Thus, educational settings have a significant influence on child SLA.

However, child learners may not benefit from being in L2-medium schools that offer limited exposure to the L2 (De Houwer, 2021: 7). Indeed, a child may resist speaking the L2 when their exposure to L2 is not as rich as their exposure to the L1 (Caldwell-Harries & MacWhinney, 2023: 7). Their limited exposure to L2 could be caused be the limited amount of input or the poor quality of this input. Thus, input quantity and quality are both significant factors in child SLA. In addition, regular exposure to two distinct languages may lead to favouring of one language over the other, depending on the quantity and quality of input in each (Unsworth, 2016b: 121). The quality of input may vary from one source to another.

Furthermore, in educational settings, teachers are likely to assist children through scaffolding language learning strategies.

Socioeconomic status is another factor that affects the quantity and quality of input to child SLA. Socioeconomic status can be measured through family income, occupational prestige and educational level, among other quantified features. Operationally, however, child language acquisition studies have often assessed socioeconomic status on the basis of maternal educational level only (Hoff, 2006: 60). Sorenson Duncan & Paradis (2020) report that mothers' level of education has a positive effect on their children's SLA. Their study involved bilingual immigrant and refugee preschool children from various L1 backgrounds. Their results show a positive impact of mothers' higher level of education on child SLA. However, they also found differences between mothers educated in their L1 and mothers educated in the L2. Immigrant children whose mothers were educated in a L2 received more linguistic input in both L1 and L2. In contrast, mothers educated in the first language mainly enriched the L1 input provided to their children (Sorenson Duncan & Paradis, 2020: 58).

Another factor that may affect child SLA is the social setting of language acquisition. For instance, homes as social settings may provide either a rich or a poor context for SLA, depending on the formal language of communications (e.g. the native language or the L2). The factors mentioned above have been found to be highly influential in shaping children's SLA processes and outcomes. It is essential to consider these factors while researching bilingual and L2 children (Paradis, 2019: 15).

2.2.2. Child language acquisition & Usage-based theory

As introduced in section 2.2.1.1, usage-based theory is an important approach within the field of child language acquisition. According to this approach, language acquisition is seen as an integrated process accomplished by means of social and cognitive skills

(Tomasello, 2003: 3). Research within this approach focuses on how children are able to construct knowledge of language through social communication (see, e.g., Tomasello, 2003; Bybee, 2010; Ellis, 2015). Proponents of this perspective argue that in the first instance, children hear and learn whole utterances (e.g. *see you later*); later, they decompose these complete utterances into their constituent parts in order to understand their functions and how to use them in other utterances (Tomasello, 2003: 40). In this model, social context is a critical factor, because children often rely on social context to understand the meaning of newly-encountered utterances. Furthermore, when communicating with adults, children always try to understand adults' communicative intentions beyond producing words or utterances (Tomasello, 2003: 75).

According to usage-based theorists, children detect patterns (i.e. new chunks of utterances) used by adults and other children; then they store and generalise these patterns (e.g. multi-word expressions) in their language use. Children become more productive once they learn how to make substitution in the parts of these learned patterns (e.g. multi-word expressions (Bybee, 2010: 35). Children employ various mechanisms to assist them in learning new language patterns. One of these mechanisms is *priming*. Priming refers to "[the] process by which a recent experience with or use of a linguistic items facilitates the processing or increases the use of the same or similar items" (Zahler, 2023: 81). For instance, if a child hears the syntactic construction "John lent Sara a pen", with a double-object, they are more likely than they otherwise would be to produce a sentence that uses the same construction, like "Ahmad bought Susan a book" or "Ahmad wrote Susan a letter", in preference to an equivalent sentence that does not use the double-object construction, such as "Ahmad wrote a letter to Susan". The study of priming from a usage-based framework focuses on two types of priming, namely, lexical and syntactic.

A number of studies report positive affects of syntactic priming on child acquisition of syntactic constructions (e.g. Huttenlocher et al., 2004; Gerard et al., 2010; Kidd, 2012; Branigan, 2016; Bidgood et al., 2020; Kumarage et al., 2022). However, some of these studies argue for the need to observe the effect of other factors that may influence the process of priming such as the speaker's linguistic representation and the age of participants. The speaker's linguistic representation refers to the abstract knowledge of the new patterns that a speaker has mentally developed before reusing these patterns. For instance, if a speaker has not formed the abstract knowledge behind the use of some specific syntactic pattern (e.g. the passive construction), their production of the primed examples is likely to be affected by their ability to reuse the lexical items or the semantic features rather than by the ability to produce a similar syntactic construction (McDonough & Trofimovich, 2015: 105). Accordingly, although priming is beneficial for the *entrenchment* of learned knowledge in child L1 and L2 acquisition, it does not play a vital role in the acquisition of new knowledge (McDonough & Trofimovich, 2015: 120; see also Gerard et al., 2010).

The other factor that affects the priming process is the age of participants. Gerard et al. (2010) find that the priming effect increases with age. The older children are, the more they are able to prime learned patterns. This finding is consistent with that of Kidd (2012), who observes the effect of priming on children's use of the passive construction in twice-repeated tests. However, some children in Kidd's study were unable to produce the primed passive constructions on the second priming test. Kidd (2012: 410) argues that due to their young age (under 7), children are still developing their working memory. This result suggests that children's acquisition could benefit from priming if it is combined with continuous reinforcement of the new patterns (Kidd, 2012: 410). How much children can learn from priming new linguistic patterns is still a matter of debate, and warrants more experimental studies (Zahler, 2023: 82).

Central to the discussion of priming within the usage-based approach is the role of *frequency* and *entrenchment*. According to usage-based theorists, constructions that are repeated/primed in language use will become entrenched patterns over time. The degree of entrenchment is closely tied to how frequently these patterns occur. Ellis (2015: 51) states that learning is often affected by "the frequency of usage: the more times we experience something, the stronger our memory for it, and the more fluently it is accessed". Consequently, more frequent patterns are often acquired earlier than less frequent ones. Within the usage-based framework, researchers often employ corpus methods[2] to examine the effect of priming and frequency of linguistic items (e.g. DMs) in children's interactions with adults and other children (Chee et al., 2023: 381). In my study, the usage-based approach is especially relevant, because I focus on studying L2 children's use of English DMs in near-natural conversations in an international school setting, meaning that the data captured consists of language usage in authentic social context. My focus on the frequency and functions of DMs in L1 and L2 children speech creates the potential for my results to served as a contribution to the empirical evidence for usage-based theory. In particular, my examination of the frequency of DMs adds to our understanding of how repeated patterns such as frequent DMs are used in children's speech. In addition, my qualitative investigation of DM functions contributes to a usage-based understanding of language development by demonstrating how children employ the forms of DMs to achieve various communicate goals. This highlights the form-function mapping central to the usage-based approach in which linguistic knowledge is shaped by repeated use of linguistic patterns in social context. I will expand on this topic in chapter 7.

[2] See section 2.3 for detailed discussion on corpus linguistics and use of corpus methods in language acquisition studies.

2.2.3. Child L2 pragmatic acquisition

Achieving competency in L2 pragmatics involves mastering sociocultural norms (Culpeper et al., 2018: 3). As we discussed earlier, children acquire these norms through interactions with adults and other children. For instance, in family interactions, adults provide models for culturally accepted conversational behaviour by initiating and supporting the communication and repairing breakdowns of interaction (Cekaite et al., 2014: 10). Through communicating with adults at home, children learn how to construct narratives and differentiate between past and future events. Children's communication with family also provides them with opportunities to learn the asymmetry of power and age with their interlocutors (Cook-Gumperz & Kyratzis, 2015: 593). However, there exist potentially restrictive effects of interacting with adults that can negatively affect a child's development of pragmatic competence, such as the asymmetry of power and knowledge between children and adults, especially teachers. Communicative tasks become easier for children with adults' assistance. Subsequently, the contributions that the children can achieve and their opportunities to practise their language skills are constrained (Cekaite et al., 2014: 10).

Children can benefit from communicating with peers – such as siblings or other children – in order to develop pragmatic skills like initiating topics, taking turns or maintaining the flow of discourse (Bryant, 2009: 351). Children utilise these skills in dyadic interactions with other children in their peer talk (Cekaite et al., 2014: 3). Peer talk has been shown to afford children opportunities for negotiating relationships in the context of L2 classrooms (Cekaite & Aronsson, 2014:212). During peer interactions, children tend to work harder to get and keep turns (Kyratzis, 2004: 641). Children help other children during playtime and peer talk to socialise and negotiate identities, especially in multilingual or multicultural settings (Goodwin & Kyratzis, 2007: 281). Peer talk in the L2 is said to enhance

learning and lead to faster acquisition (Blum-Kulka & Snow, 2004: 296). It also aids in the development of children's pragmatic and social competence in SLA (Kyratzis, 2004: 642).

Many studies have investigated the role of peer talk in the acquisition of pragmatic devices such as DMs by children in their first language (e.g. Sprott, 1992; Kyratzis & Ervin-Tripp, 1999, Andersen et al., 1999, Kim, 2011). Yeh & Huang (2016), for example, explore the L1 acquisition of the Chinese DMs *hao* "okay" and *dui* "right" in the speech of six preschool peers. They find that these children use DMs in their interactions to express equality in power relations. Other studies on DM acquisition by children show that, from a young age, children are able to use DMs to demonstrate their pragmatic competence through interactions with their peers (Kyratzis, 2004: 642). Peer talk provides a rich context to explore children's use of DMs and their level of competency in pragmatics without adult interference. Studying use of DMs in children's roleplay with peers may shed light on how children reflect on the adult modes of interaction that they have previously observed. Andersen et al. (1999) look at children's use of English DMs in three different L1 communities: English-speaking, French-speaking and Spanish-speaking. They find that children use certain DMs (e.g., *well, then* and *now*) in their native language to perform high status roles during play as a teacher or a doctor. Certain other DMs, including disfluency markers (e.g., *uh, umm* or *eh*), were used by children to perform low status roles while playing as a nurse or a student. However, the use of DMs in children's natural language interactions differs from their usage in role play. Indeed, the communicative activities within which peer talk takes place have been shown to have an impact on children's use of DMs. Using Schiffrin's (1987) model (see section 2.3), Escalera (2009) observed the functions of DMs in children's peer talk elicited in the context of various activities (including bargaining, disputes, demonstration, narration, role play, sound play, teasing and local activities). The study focuses on examining the relation between gender differences in the use of DMs and

the different activities. Escalera (2009) finds that girls used more DMs than boys did during role-play activities, but no significant difference between genders were observed during other activity types. Odato (2013) reports similar results from a follow-up study on acquisition of *like* as a DM by children of different ages and genders, finding that girls used DM *like* more frequently than do boys.

Peer communication can also play a key role in language change and in the transmission of vernacular rules that might not be acquired at home. For example, the use of be+*like* as a quotative, instead of reporting verbs such as *say* or *go*, developed in adolescence speech and spread mostly among peers (Levey, 2006: 173; see also Andersen, 2001; Beeching, 2016; D'Arcy, 2017). Andersen's (2015) study lends additional weight to the importance of peer talk. Anderson reports that young children acquiring Norwegian begin to use the verb *se* 'see' to fulfil discourse functions that differ from its uses in adult Norwegian. This non-adult-modelled behaviour, Andersen argues, emerges from children communicating with peers in schools and preschools. Such observation makes it highly interesting to investigate children's pragmatic competence and peer interactions to examine the influence that they have in the emergence of new pragmatic strategies in L1 and L2 acquisition. Previous studies on DM acquisition in children have considered only L1 contexts. In this study, I offer an additional perspective by looking at this issue in SLA contexts.

2.3. Discourse markers

2.3.1. Definition and terminology

DMs are a heterogeneous class of linguistic expressions which operate as signals of the relationship between units of discourse, guide the listener to the upcoming utterance (Redeker, 1990: 372) and similarly signal " an interactive relationship between speaker, hearer and the message" (Fung & Carter, 2007: 411). Although numerous studies observe

these markers in different contexts, a universal identified set and definition are still debatable. The challenge in defining DMs comes from the different approaches adopted in examining them (Schourup, 1999: 228); for instance, DMs under the field of discourse analysis contribute in establishing coherence relations between discourse segments (Schourup, 1999: 244). This understanding has prompted some to adopt the term *discourse markers*, when discussing coherence relations (e.g., Schiffrin, 1985, 1987, 2001, 2006; Fraser, 1988, 1990, 1999; Jucker & Ziv, 1998; Brinton, 1996, 2006; Blackmore, 2006; Müller, 2005; Fung & Carter, 2007; Furkó, 2020; Crible, 2017, 2018).

Schiffrin (1987) presents an influential model of DMs under Coherence Theory, investigating their functions in constructing coherence relations between discourse units. From this perspective, DMs are "sequentially dependent elements which bracket units of talk" (Schiffrin, 1987: 33). The canonical definition is elaborated later to embrace the function of DMs as contextual coordinates in discourse (see Maschler & Schiffrin, 2015). DMs in this sense are employed to link an utterance with prior discourse to achieve coherence (Schiffrin, 1987: 326). Investigations of DMs according to Schiffrin's preliminary definition incorporate analyses of eleven elements from different grammatical classes, varying from conjunctions (*and*, *but*, *or*, *because*) to interjections (*oh*), adverbs (*now*, *then*) and lexicalised phrases (*you know*, *I mean*).

Another term for DMs used by researchers is *discourse particles* (e.g., Schouroup, 1983; Mosegaard Hansen, 1998; Fischer, 2000, 2006, Aijmer, 2002). This label, however, is deemed "problematic" (Schourup, 1999: 229) since it implies a syntactic category, when the associated items do not fit into any single syntactic class (Zwicky, 1985: 292). Rather, as the aforementioned list of eleven elements suggests, the class of functional words defined to be DMs is derived from syntactically heterogenous groups (Furkó, 2009: 19). Hence, the term *particle* is mostly used to refer to the group of DMs in other languages such as German

(Müller, 2005) or Chinese (Mei, 2012). One solution to the problem of DMs' syntactic heterogeneity is to note, following Fischer (2006: 5), that the term DM is regarded as a functional category more than a formal category of a word class or a particle.

The terms *discourse operators* (Redeker, 1990, 1991) and *connective*s (Celle & Huart, 2007) have also been applied to refer to coherence elements of discourse. Redeker (2006: 340) argues that DMs are a class of *discourse operators*. Redeker (1991: 1164) defines a *discourse operator* as a spoken word or phrase whose primary function is to bring the listener's attention to a link from the upcoming utterance to the previous discourse context (Redeker, 1991:1168). The implementation of the term *connectives* (Celle & Huart, 2007) and *discourse connectives* (Blackmore, 2006) also limits the class of DMs to include only connectors between discourse segments. These terms restrict DMs to signal only coherence relations.

The study of DMs, however, is not restricted to the analysis of discourse only. Several studies have investigated DMs form the perspective of pragmatic theories. Levinson (1983: 85) refers to a class of words and phrases that others consider DMs under the pragmatic label of "discourse deixis", which indicates the relation of an utterance to adjacent text. Levinson (1983: 87), however, does not assign a name to this group of words and phrases. In his account, these elements work as continuation signals that link their host utterance with prior ones. The meaning of these elements is often constrained through conversational implicature (Levinson, 1983:100).

One of the other most widely used terms for the group of DMs is *pragmatic markers* (henceforth PMs). Many researchers use the term PMs as an equivalent to or substitute for *discourse markers* (e.g. Andersen, 2001; Erman, 2001; Brinton, 2017). However, for other scholars, PM is an umbrella term for a wider set of markers which includes DMs (e.g. Fraser,

1996; Aijmer & Simon-Vandenbergen, 2006; Mosegaard Hansen, 2006). For instance, Fraser (2009: 307) classifies DMs as a type of PMs. Fraser (1999: 950) argues that a lexical expression functions as a DM if it links two meaningful discourse segments. The DM often occurs at the beginning of the second segment and signals a semantic relationship between discourse segments (Fraser, 2006: 191).

Östman (1981: 5) discusses DMs as *pragmatic devices* that implicitly link a speaker's attitude towards an utterance in relation with the current interaction. Östman (1981: 5) also labels them *pragmatic expressions*, *phrases* and *particles*. In Östman's account, speakers use PMs to reflect two pragmatic concepts: face-saving and politeness. These concepts are tied to speakers' and listeners' roles in keeping track of the interaction. In this sense, speakers often try to control and modify their speech to meet their listeners' communicative needs. PMs are also linked to contextual factors, such as the attitude of speakers towards the uttered proposition. They facilitate pragmatic processes by constraining the intended meaning (Andersen, 2001: 22) as well as contributing to face-saving and politeness during communication. Östman (1981: 4) argues that the absence of PMs results in viewing speakers as rude or impolite.

The study of DMs has been tackled not only from pragmatic and discourse perspectives, but also in terms of sociolinguistics and variational pragmatics (e.g. Andersen, 2001, Aijmer, 2013; Beeching, 2015, 2016). Aijmer (2013: 4) discusses the role PMs (in Aijmer's terms) play in communication in relation to two sociolinguistic concepts: *contextualisation* and *metalinguistic indicators*. DMs work as contextualisation cues in the interaction or conversation. They indicate transition in social context, and facilitate hearers' understanding of speakers' intentions in the ongoing interaction. Aijmer (2013: 4) emphasises that speakers' cognitive processes are hidden to others, but can be mirrored out in conversation with DMs, which function as metalinguistic indicators. Thus, DMs are reflexive

indicators of "metapragmatic awareness" of speakers' choices about what and how to say something (Aijmer, 2013: 4). DMs in Aijmer's (2013: 5) account aid in organising the discourse and in achieving coherence.

To summarise, most scholars have agreed on the view that DMs are a class of linguistic expressions that provide connectivity, aid in organising the discourse and mark the speaker's relation to the message and hearer(s) (Furkó, 2020: 1). Furthermore, most definitions of DMs agree on some features that a word or a phrase must possess to be categorised as a DM. In the present study, the term *discourse markers* is adopted as the most general term (Fischer, 2006: 6) and one that encompasses alternative labels discussed in this section. According to Jucker & Smith (1998: 2), DMs is a term that "enables us to include a broad variety of elements under a single conceptual umbrella". In this sense, the term DMs is a practical cover term and not a strictly defined term that derives from a specific theoretical approach.

2.3.2 Theories of discourse markers

The definitions of DMs and their theoretical assumptions vary according to the perspective of the different linguistic fields that are used to describe them (e.g. semantic, pragmatic or discourse analysis). Among these approaches, three merit detailed consideration due to their profound influence on DM studies in the last few decades.

2.3.2.1. Discourse markers & Coherence-Based Theory

The first set of perspectives on DMs that I will discuss derive from Coherence-Based theory. From a coherence-based account, DMs aid in identifying coherence relations between adjacent textual units (Schourup, 1999: 240). Two models of coherence are employed in the study of DMs: Schiffrin's (1987) model, known as the *integrative model*; and Fraser's (1988) model, known as the *grammatical-pragmatic* model.

Schiffrin's (1987: 22) integrative model of coherence is grounded in the contribution of DMs' properties to overall discourse coherence (Schiffrin, 1987: 22), in the sense of local coherence between adjacent utterances. Three properties of discourse related to coherence have been identified: structure, proposition and action (Schiffrin, 1987: 7). According to Schiffrin's (1987: 22-27) integrative model, coherence is fulfilled in one of five planes of talk: a) an *action structure* that marks a speech act; b) an *exchange structure* which denotes the turns of talk, a sequence of linked utterances; c) an *ideational structure* in which ideas form the main cohesive tools; d) a *participation framework* for the relationship between speaker, hearer and the talk; and e) an *information state* which focuses on knowledge and meta-knowledge management on the part of the speaker.

Redeker (1991: 1139) criticizes Schiffrin's (1987: 22-28) model for lacking "clarity and consistency in the definitions and the use of theoretical terms and analytical categories". In a review of Schiffrin (1987), Redeker (1991: 1162) argues that the "participation framework" and "information state" should not be considered to be among the planes of talk because they are unrelated to inter-utterance coherence; rather, they mainly relate to interpretation of single utterances. Redeker (1991: 1169) proposes a revised coherence model which consists of three main planes: *ideational structure*, which identifies the relations between discourse segments (e.g. temporal, casual, etc.); *rhetorical structure*, which expresses the speaker's intentions; and *sequential structure*, which signifies the "paratactic or hypotactic" relations between discourse units. While the first two of these planes correspond with planes of talk in Schiffrin's (1987) model – namely ideational and action structure – the sequential structure is an extended account of Schiffrin's (1987) exchange structure. According to Redeker's model, the primary function of a DM (discourse operator in Redeker's terms) is to bring "the listener's attention [to] a particular kind of linkage of the upcoming utterance with the immediate discourse context" (Redeker, 1991: 1169).

Fraser (1988, 1990, 1996, 1999, 2009 and 2021) suggests a Grammatical-Pragmatic model to study DMs as markers of coherence. While Schiffrin (1987) argues that DMs link adjacent units of talk, creating *local* coherence, Fraser (1999) notes that a DM can link a segment of discourse with any part of the previous discourse. This constitutes *global* coherence. Moreover, Schiffrin (1987) focuses on the structural role of DMs in building coherence, whereas Fraser's account (1999) concentrates on the cognitive role of DMs in establishing coherence. Fraser (1988:21) notes that *pragmatic* meaning is manifested through the use of *pragmatic markers.* Unlike the aforementioned discussion of the term PMs (see section 2.3.1), Fraser (1996) uses PMs as a general term that encompasses different types of markers. Fraser (1996) states that these markers are:

- *Basic markers:* signalling the force of the basic message that is being communicated. Examples include *I regret* and *admittedly*.
- *Commentary markers:* showing the speaker's attitude towards an action or state via comments on the basic message. Examples include *frankly, amazingly, reportedly, fortunately, certainly* and *sadly.*
- *Parallel markers:* expressing a separate message from the basic or commentary message. Examples include *sir*, *your honour, now, well* and *ok.*
- *Discourse markers:* indicating the relationship of the discourse segments where they occur to the prior discourse. In earlier work (Fraser, 1988: 22; Fraser 1990: 387), Fraser considers these a category of *commentary pragmatic markers*, but as of 1996, Fraser treats *discourse markers* as an independent type of pragmatic marker.

According to Fraser (1990), DMs have core meanings that are not constructed from the content meaning of a sentence. Rather, Fraser deems the meaning of a DM to be 'procedural', not 'conceptual'. While conceptual expressions do not require prior context to clarify their meaning, DMs are procedural expressions: their explanation entails looking at

prior discourse segments (Fraser, 1999: 945). However, in later work, Fraser (2006: 31) revises this view and agrees with one of the Relevance-based proponent's claim (see. Blakemore, 1992) that DMs can encode both meanings.

2.3.2.2. Discourse markers & Relevance-based Theory

The second set of perspectives is based on Relevance Theory, proposed by Sperber & Wilson (1986): a "theory of pragmatics" that focuses on identifying the *principle of relevance* (Blakemore, 1992: 36). This principle distinguishes two sides of relevance: *cognitive* and *communicative* (Wilson & Sperber, 2002: 249). The cognitive principle entails that "human cognition is geared to the maximization of relevance" (Wilson & Sperber, 2002: 254). The communicative principle is that any single utterance carries an expectation of "optimal relevance" (Wilson & Sperber, 2002: 256).

Blakemore (1992, 2002) proposes a relevance-based model of DMs (*discourse connectives* in Blakemore's model). According to the Principle of Relevance, an utterance is relevant if it generates a suitable cognitive effect from a speaker with the least processing cost or effort from a hearer (Andersen, 2001: 309). In this sense, a speaker works to constrain the context so that the intended message is readily interpreted by the hearer (Blakemore, 2002:70). The hearer perceives the discourse as coherent when they can establish relevance between the utterance and the prior context (Jucker, 1993: 440). Thus, in Relevance Theory, coherence depends on recognising the optimal relevance of each utterance to the upcoming one (Jucker, 1993: 440).

According to the relevance model, there should be cooperation between the speaker and the hearer to achieve optimal relevance (Blakemore, 2002:64). A hearer utilises the Principle of Relevance by considering all the different interpretations of an utterance. Therefore, the hearer will opt for the single most suitable utterance to the intended message

of the speaker (Andersen, 2001: 14). The speaker's role is to constrain the context so that the intended message is readily interpreted by the hearer (Blakemore, 2002:70). Blakemore (1992: 138-142) identifies four types of DMs that prove the information supplied in the utterance to be relevantly depicted through using DMs. DMs can either introduce contextual implication (e.g., *therefore, so*); strengthen an existing assumption by evidence (e.g., *after all, moreover, furthermore*); introduce denial (e.g. *however, but*); or adds further information (e.g. *also*) (Blakemore, 1992: 138-142).

A significant issue raised by Blakemore in accounting for the meaning of DMs is the distinction between two types of semantic meanings: *conceptual* and *procedural*. Conceptual meaning contributes to the proposition of an utterance and determines its truth conditions. This type of meaning is explicated in lexical expressions (e.g. nouns, verbs, adverbs and adjectives). Procedural meaning, however, has no effect on the truth condition of the utterance, and is often encoded in grammatical expressions (Wilson, 2011: 2). For Blakemore (1992, 2002), DMs encode procedural meaning only. However, the conceptual/procedural distinction, with regards to DMs, should be seen as complementary feature to distinguish DMs rather than as clear cut (Pons Bordería, 2008; Fraser, 2006).

Studies of DMs based on Relevance Theory have been criticized for ignoring the diachronic use of DMs as well as their occurrence in different text types (Lam, 2007: 354). Aijmer (2013: 11) notes that studying DMs needs investigating principles that are less general than relevance, including specific cultural or social situations, among other particular contexts.

2.3.2.3. Discourse markers & Meaning Potentials Theory

A contemporary model that accounts for DMs' meanings has been proposed by Aijmer (2013, 2015, and 2016). The model is based on the early semantic theory of Meaning

Potentials, in which a *meaning potential* refers to all the meanings a word might possibly have, depending on context. Then, in any given context it has just one of those meanings. As Allwood (2003:16) explains, the meaning potential refers to "all the information that the word has been used to convey either by a single individual or, on the social level, by the language community". The purpose of using *meaning potential* is to specify the intents of both speakers and hearers to "mean and understand things in context" (Norén & Linell, 2007: 387). Meaning potentials include both lexical and encyclopaedic information (Allwood, 2003: 16).

Aijmer (2013: 12) applies the theory of Meaning Potentials to DMs (pragmatic markers in Aijmer's account). The theory is compatible with the polysemy of DMs in the sense that both notions look at the core meaning of a marker and possible interpretations of that marker (Aijmer, 2013: 12). The Meaning Potential theory accounts not only for the notion of DMs having several conventionalised meanings but also for the idea that meanings can conform to new situations (Aijmer, 2015:202). The creativeness of the theory makes it possible to explain the innovative use of markers in specific contexts (Aijmer, 2013: 149). The significance of the theory of meaning potentials lies in its ability to account for features of DMs such as forms, positions and functions (Aijmer, 2015:202). Contrasting this approach to Relevance Theory, Aijmer (2013: 12) argues that in Meaning Potential Theory, "the context selects the meaning of a pragmatic marker", whereas in relevance theory, "a pragmatic marker looks for a context which is compatible with communicative principles". The selection of a meaning promotes the need for describing contextual factors that integrate with the meaning potentials (Aijmer, 2013:12).

Let us compare these different theoretical approaches to DMs. Coherence supporters consider DMs to be linguistic expressions that encode cohesion relationships between

discourse units. Thus, DMs contribute to the coherence of discourse. According to Fraser's model, DMs carry a core meaning which is negotiable according to the context (Fraser, 1999: 950). By contrast, relevance theorists believe that DMs encode procedural meaning, which constrains the choice of utterance meaning and accordingly governs relevance relations between discourse units. Coherence in Relevance Theory is achieved through the hearer's use of information as they strive to determine relevance. Aijmer's model focuses on how to investigate the potential or polysemous meanings of DMs according to different factors such as text types, activity types, speakers and degree of the speech formality (Aijmer, 2013: 2).

All the three discussed theories attempt to explain the role of DMs in linking discourse segments and marking the relationship between speakers and hearers.

2.3.3. Characteristics of discourse markers

Despite differing theoretical stances, there is some consensus in the literature on defining characterises of DMs as distinguishable from other linguistic elements. These characteristics are related to the notion of DMs as important cues that signal relations between segments and guide the listener's understanding of the discourse (Furkó, 2020: 7). Östman (1982: 149) observes a set of structural criteria that delineates a universal characterisation of DMs:

i) a DM is characterised as short and prosodically subordinated,

ii) it does not contribute to the propositional meaning of a sentence,

iii) it occurs outside the sentence structure.

Many researchers have listed other characteristics of DMs (e.g. Schiffrin, 1987; Schourup, 1999; Aijmer, 2002, 2013). There have been also various attempt to summarise these characteristics. The summary favoured in this study is that of Brinton (1996, 2017 and

2023). Brinton's earlier list (1996: 33-35) includes most of the features discussed by Östman (1982) and Schourup (1999). Recently, Brinton (2017: 9, 2023) modified the list of DM characteristics, categorising DMs into five linguistic domains, as presented in table 2.1. Some of these characteristics are non-essential, i.e. the *sociolinguistic and stylistic characteristics*. The discussion of DMs in this study will therefore be based on the four other domains: *phonological and lexical*, *syntactic characteristics*, *semantic characteristics* and *functional characteristics.*

Table 2.1. Characteristics of discourse markers (adapted from Brinton, 2017: 9*).*

Phonological and lexical characteristics
(a) Pragmatic markers are often "small" items, although they may also be phrasal or clausal; they are sometimes phonologically reduced.
(b) Pragmatic markers may form a separate tone group, but they may also form a prosodic unit with preceding or following material.
(c) Pragmatic markers do not constitute a traditional word class, but are most closely aligned to adverbs, conjunctions, or interjections.
Syntactic characteristics
(d) Pragmatic markers occur either outside the syntactic structure or loosely attached to it.
(e) Pragmatic markers occur preferentially at clause boundaries (initial/ final) but are generally movable and may occur in sentence- medial position as well.
(f) Pragmatic markers are grammatically optional but at the same time serve important pragmatic functions (and are, in a sense, pragmatically non- optional).
Semantic characteristics
(g) Pragmatic markers have little or no propositional/ conceptual meaning, but are procedural and non-compositional.
Functional characteristics
(h) Pragmatic markers are often multifunctional, having a range of pragmatic functions.
Sociolinguistic and stylistic characteristics
(i) Pragmatic markers are predominantly a characteristic of oral rather than written discourse; spoken and written pragmatic markers may differ in form and function.
(j) Pragmatic markers are frequent and salient in oral discourse.
(k) Pragmatic markers are stylistically stigmatised and negatively evaluated, especially in written or formal discourse.
(l) Pragmatic markers may be used in different ways and in different frequencies by men and women

Brinton's list is based on Schourup's (1999) seven characteristics of DMs: *connectivity*, *optionality*, *non-truth conditionality*, *weak clause association*, *initiality*, *orality and multi-categoriality.* But unlike Schourup, Brinton (2017) offers a detailed,

comprehensive list of DM characteristics across multiple linguistic fields, i.e., phonology, syntax, semantics, pragmatics and sociolinguistics. Thus, Brinton's list is an elaborated version of Schourup's. For example, two of Brinton's criteria, (b) and (d), correspond to Schourup's (1999: 230) assertion of 'weak clause association' – that is, that DMs are either "loosely attached or outside the syntactic structure" and consequently are prosodically independent elements (Furkó, 2020: 10). In examining the functions of *well* and *so* as DMs, Lam (2007: 270) argues that a characteristic that distinguishes these two forms from their non-DM counterparts is their high degree of prosodic autonomy as DMs. However, despite their prosodically and syntactically detached nature, we cannot conclude that DMs are *completely* outside the syntactic structure (Lewis 2006: 44). The fact, already discussed, that DMs come from different syntactic word classes is on Brinton's (2017) list as criterion (e), and Schourup (1999) labels this *multi-categoriality*. But for Lewis, this feature shows that DMs have their own position in syntactic theory, forming a separate group due to their functions at the discourse and pragmatic level (Lewis, 2006: 44).

Schourup (1999) highlights 'initiality', that is, the characteristic that DMs mainly appear in utterance initial position. Fraser (1999: 950) argues that a DM should be the initial element in the host segment to show its connectivity to prior segments. However, many studies have proven that DMs can occur in medial or final position (Biber et al, 1999: 1086, see also Aijmer, 2002, 2013; Crible 2017; Brinton, 2017). Brinton (2017) asserts that DMs preferentially occur at clause boundaries, but can also occur in sentence-medial positions (criterion (e)). Indeed, many studies find that DMs are mobile items that occur in different positions. Their mobility is one of the reasons for their multifunctionality. Beeching and Detges (2014: 13) argue that a DM in initial position, also known as the left periphery, often serves a discourse-structuring function such as topic initiation or topic change. But if the DM

appears in final position, i.e. right periphery, it often signals interpersonal functions such as relinquishing the floor or initiating a re-negotiation of the topic (see section 2.3.4).

One of the defining characteristics in DMs is their optionality. Brinton (2017: 9) illustrates that this criterion defines DMs from a syntactic perspective. In her view, a DM is grammatically optional but "pragmatically non-optional", as it serves a significant pragmatic function (criterion (f)). On the other hand, Schourup (1999: 231) describes DMs as optional items from two perspectives. First, DMs are optional items at the level of syntactic relationships among sentence parts. Hence, omitting a DM will not change the grammar of its host sentence (Schourup, 1999: 231), though it does remove an important clue about the speaker's commitment to the relationship between current and previous discourse (Fraser, 1988: 22). The other perspective on optionality is that omitting a DM will not affect the relations between utterances, as its meaning is still conveyed to the hearer (Schourup, 1999: 231). However, deleting a DM may result in changing the different possible interpretations of the discourse (Mosegaard Hansen, 2006: 26). Characterising DMs as optional means that using these markers does not change the sentence well-formedness, whereas omitting them can still change the pragmatic effect of an utterance (Müller, 2005: 4).

Brinton (2017:9) defines DMs as having "little or no propositional/conceptual meaning" (criterion (g)); this corresponds to Schourup's (1999) argument that they are *non-truth conditional* elements. Some researchers explain the non-truth conditionality of DMs as their not affecting the sentence meaning (e.g. Fraser, 1996: 167). Others believe that non-truth conditionality does not change the truth of the mental representation of a sentence (e.g. Blackmore, 1988: 16). Researchers who argue for DMs as truth conditional elements often include the class of adverbials such as *frankly*, *indeed*, and *I think* on the list of DMs (e.g. Furkó, 2009; 2020). Furkó (2009: 56) argues that the distinction of DMs as truth conditional or non-truth conditional elements should be viewed as a continuum and not as a dichotomy.

This is because DMs are undergoing a process of grammaticalisation in which these elements change their syntactic categories[3]. This view is compatible with observations of the functions of the DM *like*. Siegel (2002) differentiates the functions of *like* as a DM (*discourse particle* in Siegel's account) according to their effect on the truth condition of the utterance. When *like* has an approximative function (e.g. John watched *like* three movies), Siegel believes that its omission affects the truth condition of the utterance. *Like* in this example expresses that the speaker is not sure whether John watched precisely three movies or more/less than that. Omitting *like* would make the speaker sound more confident that the number is precisely three. It shows that a DM can carry a "little" conceptual meaning that may affect the truth conditionality of its utterance.

Schourup (1999) positions DMs as features of spoken discourse mainly; likewise Brinton (2017) argues that DMs "are predominantly a characteristic of oral rather than written discourse". Orality being treated as a defining feature of DMs is a result of the many studies that have focused on DMs in speech data (Schourup, 1999: 234), and/or of their relatively high frequency in speech data (e.g., Beeching, 2016). But DMs are not limited to speech; spontaneous written discourse (e.g., online debate or instant messaging) proves to be rich in DMs (Brinton, 2017: 4). Still, whereas in speech DMs function both to signal coherence in discourse and to display interpersonal relations (Brinton, 2017: 3), their role in written discourse may be limited to creating coherence relations. While many researchers emphasise DM orality as a compulsory feature, Brinton (2017: 5) describes this as a tendency and not a requirement. I will follow Brinton's view of this characteristic as non-essential.

[3] There is a long debate in the literature on the grammaticalisation process of DMs from a diachronic view (see Brinton, 1996, 2006; Traugott, 1995; Romaine & Lange, 1991; D'Arcy, 2017 among others).However, the developmental path of DMs is beyond the scope of the current study.

2.3.4. Functions of discourse markers

A defining characteristic of DMs in Brinton's list and many other researchers' is their multifunctionality (criterion (h) in table 2.1). Although most researchers agree that DMs are multifunctional, their analyses of DM functions vary according to the theoretical framework that they adopt (coherence approach, relevance approach, etc.). Schiffrin (1987) explains the functions of DMs according to her concept of five planes of talk (section 2.3.2.1). A DM can operate in multiple planes, having a primary and a secondary function. For relevance theoretical researchers, a DM's function is to guide the hearer to the most suitable explanation of an utterance (see Blakemore, 1987, 2002). Thus, the function of a particular DM can vary across contexts. For instance, Jucker (1993) argues that the DM *well*, explained from a relevance theoretical view, can carry four different functions according to its context: a signal of insufficiency (marking some problem in the upcoming utterance), a face threat mitigator, an introduction of a new topic, or a delay marker (see further section 4.5.1). By using *well* with these various functions, speakers direct their hearers as to how they should process the uttered speech (Jucker, 1993: 438).

Many researchers seek to identify the various functions of DMs through the model of cohesion suggested by Halliday and Hassan (1976). This model identifies three metafunctions of language: *ideational, textual* and *interpersonal*. Ideational functions are related to the semantic meaning or the content of expressions, including logical and experiential relations (Halliday & Hassan, 1976: 26). While textual functions mark the structuring of a discourse; interpersonal functions signal the speaker's attitude and the management of social exchange (Brinton, 2006: 310). Following Halliday & Hassan's (1976: 26) metafunctions categorisation, many researchers propose various models of functions to account for the multifunctionality of DMs. Some employ three levels of functional

categorisation (e.g. Fung & Carter, 2007; Buysse, 2012; Crible & Zufferey, 2015; Bakeer, 2023). Besides adopting Halliday and Hassan's (1976) model, some of these studies follow Redeker's (1991) distinction between rhetorical and sequential types of DMs instead of grouping them under one broad domain, i.e. textual. Other researchers propose a two-fold account for the functions of DMs: textual and interpersonal (Brinton, 1996, 2006, 2017; Andersen, 2001; Aijmer, 2002, 2013; Castro, 2009; Müller, 2005; Buysee, 2007; Ament & Parés, 2018 among others), omitting ideational functions (i.e. the expression of propositional content). In terms of their textual functions, DMs are used to:

- initiate and close discourse
- aid the speaker in controlling the floor
- work as filler or delaying strategy
- mark discourse boundary (e.g., a new topic, topic shift)
- indicate either old or new information
- repair the speaker's discourse or the discourse of hearers

(Summarised from Brinton 2006: 310)

The interpersonal domain of DMs, according to Brinton (2006: 310), consists of two functions. First, DMs may act as a reaction or a reply to the prior discourse or express an attitude to the upcoming discourse. The other interpersonal function of DMs is to establish cooperation and intimacy between the speaker and the hearer (e.g., confirmation of shared knowledge, checking understanding, saving face) (Brinton, 2006: 310). Brinton (1996: 39) uses this classification in an analysis of historical texts and the evidence they provide for the diachronic development of DMs. However, defining the boundaries between the two domains of functions is not always straightforward, as shown in Brinton (1996, 2006, and 2017). For

instance, the function of "holding the floor" is classified as textual, but it also has an interpersonal dimension.

Aijmer (2002: 40) offers another explanation for the textual and interpersonal functions. Within the *textual* domain, Aijmer (2002), following Lenk (1998), describes DMs as functioning at either a local or global level of coherence (see. section 2.3.2.1). At the global level, DMs have *frame* functions, and at the local level, they function as *qualifiers* (Aijmer, 2002: 40). The *frame* functions refer to the use of DMs when there is a need to bring the hearer's attention to a break or a shift in the conversation. The frame function includes subfunctions such as a) introducing a new turn, topic or direct speech, b) marking transition in discourse, and c) signalling self-correction. On the other hand, a DM functioning as a *qualifier* signals the need for modification when the conversation proceeds in the wrong direction (Aijmer, 2002:46). This function often appears before a question, or before an argument. Aijmer (2002: 48) reconceptualises the *interpersonal* functions as *phatic* functions. Phatic functions are the expression of attitudes and emotion. While phatic function is associated mainly with politeness and face-saving, it also fulfils other subfunctions related to planning processes (e.g. *sort of*). DMs can also express phatic functions "to request attention (*you see*), to assume shared knowledge (*you know*), to request confirmation (*right, OK*) or to clarify something (*I mean*)" (Aijmer, 2002: 52).

In sum, different functions of DMs (e.g. addition, contrast, monitoring, cause and so on) are discussed under different domains (e.g. textual, interpersonal, rhetorical or ideational) in the literature (Crible & Degand, 2019: 12). Some of the variation in these domains and functions arises from differences in the theories and approaches utilised. Nevertheless, we may safely say that the functions of DMs remain contested among researchers in the field. For present purposes, DMs are defined as linguistic expressions whose main functions are to connect parts of the discourse at local and global level and to aid in structuring the discourse

flow (Crible & Blackwell, 2020: 25), alongside carrying interpersonal functions. Consistent with previous research on DMs, I will discuss the various functions of DMs in terms of the textual and interpersonal macro-functions; my proposed framework of DM functions will be introduced in section 3.5.2.

2.3.5. Discourse markers and second language acquisition

In the field of L2, many studies have investigated the use of one or a set of markers by adult L2 speakers of English with different language backgrounds. For instance, Müller (2005) compares L1 and L2 speakers' use of DMs, looking at four DMs (*well*, *you know*, *like* and *so*) in the speech of German learners of English compared to American English speakers. The significance of Müller's (2005) study is that it offers a detailed description of the various DM functions in the speech of both L1 and L2 speakers. However, although Müller (2005) points out differences between the two groups in her analysis, she does not explain the absence of some functions in the speech of L2 speakers. House (2013) argues that although German learners of English's use of DMs differs from that of L1 speakers, they are able to use DMs to serve various connective functions that prove their pragmatic competence in their L2. According to House (2013), such differences between L1 and L2 speakers show that learners are creative in their use of DMs to serve their own communicative purposes.

Researchers have also explored the use of DMs by L2 learners in pedagogic settings. For instance, Fung and Carter (2007) find that L2 Chinese learners use DMs less frequently than English speakers do. Moreover, learners often utilise DMs to perform textual and cognitive functions rather than interpersonal functions. In a similar language context, Huo (2023) compares the use of five DMs (*well*, *like*, *you know*, *I mean* and *but*) in a corpus of Chinese learners of English speech with another corpus of English speakers. The study employs a mixed research method, using corpus tools for quantitative analysis and manual

inspection for more in-depth qualitative analysis. The quantitative results show that L2 learners use the five DMs less frequently than the English speakers do. *Well* is the most used DM of these five in both corpora, and *I mean* the least used. Through qualitative analysis, Huo (2023) finds that the Chinese learners use these five DMs to serve textual functions, whereas the English speakers employ them mainly for interpersonal functions. Huo (2023) argues that the underuse of interpersonal DMs in the speech of Chinese learners is due to the lack of pedagogic implementation of these markers in classrooms. Teachers often focus on teaching goals that do not primarily involve introducing DMs (Huo, 2023: 174). Thus, there is always a need for teachers to include interactive tasks (e.g. group task, peer conversations) to provide a good opportunity for learners to practice a wide variety of DMs.

Alongside the formal context of study, the amount of language input can significantly affect the acquisition of DMs. Ament & Pérez (2018) compare the use of English DMs by Spanish undergraduates grouped according to the amount of English input they receive, with one group studying a curriculum of English-medium courses and the other studying only one course in English. The results show that the English-medium group are more likely to use DMs to signal interpersonal and cognitive functions, while the group studying a single English course only use DMs to mark referential, i.e. textual, functions. Ament and Pérez (2018) observe that most of the referential use of DMs in L2 speech is linked to the grammar classes the speakers received during their study. Thus, the formal setting in which learners encounter DMs might have a great effect on their use of DMs.

Some studies of DMs find that the acquisition of DMs is often linked to naturalistic exposure rather than formal settings. One such study is Bu (2013), who observes frequencies and functions of DMs in interviews and classroom discussion by Chinese learners of English. Bu finds that learners use DMs less frequently than English speakers. In addition, learners employ DMs more frequently in interviews than in classroom discussion. Furthermore, the

use of certain DMs varies according to the context of discussion. For instance, in classroom discussion, learners tend to use *right* and *okay* more frequently, whereas in interviews, they prefer to use the DM *yeah*. Liu (2016) reports similar results in terms of less frequent use of DMs by Chinese learners, in comparison to English speakers. However, Liu's study shows that learners who have greater exposure to L2 in English-speaking countries use DMs more frequently than learners who spend less time in English-speaking countries. These studies are in line with previous studies that report higher frequencies of DMs in the speech of Chinese learners with some exposure to English in naturalistic settings (e.g. Hellermann & Vergun, 2007; Liao, 2009).

Gilquin (2016) observes the use of DMs in three different corpora to explore the influence of naturalistic input on the acquisition of DMs by L1 and L2 speakers. The corpora were: the Louvain International Database of Spoken English Interlanguage (LINDSEI; Gilquin et al., 2010), the Louvain Corpus of Native English Conversation (LOCNEC; De Cook, 2004) and the International Corpus of English (ICE; Greenbaum & Nelson, 1996). While the first corpus consists of spoken interviews with advanced learners of English from different L1 backgrounds, the second is a counterpart corpus covering interviews with native speakers of English. The last corpus contains speech and written data from speakers of various English varieties (e.g. Indian English, Hong Kong English and Philippine English). This study shows that more exposure to L2 has a positive effect on the acquisition of DMs (Gilquin, 2016: 244). However, the classroom setting was not conducive to the acquisition of DMs, as was also discovered in previous studies (e.g. Müller, 2005; Gilquin, 2008). Gilquin's (2016) study also demonstrates that L2 speakers employ *like*, *sort of* and *you know* as DMs more than learners with more limited exposure to English (or foreign language learners, in Gilquin's term). Gilquin (2016: 235) argues that these markers are more common in the

speech of those who have had natural exposure to the L2, that is, that natural exposure to L2 facilitates "near-native" appropriate use of DMs (Gilquin, 2016: 243).

The above-mentioned studies focus on the use of DMs in the speech of adult language learners. Studies into child SLA of DMs are more limited. For instance, Romero-Trillo (2002) studies the use of English DMs by Spanish-speaking children and adults compared to English-speaking child and adult usage. Romero-Trillo (2002) observes *Pragmatic Fossilisation* in L2 speech, that is, the outcome of inappropriate use of forms at the pragmatic level (Romero-Trillo, 2002: 770). While L2 children fossilise in their use of some DMs, i.e. *listen* and *look*, adults avoid using them. Unlike L1 children, L2 children use the DM *listen* more frequently than the DM *look*; the translation equivalent of the latter is seen as politer than the equivalent of *listen* in Spanish. Romero-Trillo (2002) argues that these children's reliance on DM *listen* over *look* is a result of a lack of naturalistic pragmatic input in formal educational settings.

Jakupčević (2019) observes the use of DMs in two groups of Croatian children learning English as an L2 (ages 6-7 and 10-11 years) in primary school, collecting data by recording the narration of a picture-book story and via a self-report questionnaire, and comparing results to Fung & Carter's (2007) findings on L1 speakers. Jakupčević finds that the Croatian children use DMs less frequently than the L1 speakers. Qualitative analysis shows that the most prevalent DM in the data, *and*, functions as a signal of the sequence of events in the narrative, as a transition, and as a primary discourse management tool (Jakupčević, 2019: 420). Other low-frequency DMs include *then, because* and *but.* The DMs in Jakupčević's (2019) data are infrequent due to the use of narrative as the only method to elicit DM usage. Moreover, the children in Jakupčević's (2019) study had had limited exposure to English as a L2, four short sessions a week only. Nevertheless, Jakupčević (2019) enhances our understanding of how children utilise DMs to mark relations between

segments of the narrative, keeping their turn and gaining time to think and develop their ideas. This study also illustrates positive development of children's L2 pragmatic competence despite limited exposure to L2.

The review of studies of DMs in the speech of L2 adults and children has highlighted some differences in how learners use DMs compared with L1 speakers. These differences may be indicative of aspects of L2 learners' pragmatic competence. They also enhance our understanding of how the context of learning the L2 can affect use of DMs. For instance, underuse of DMs in children's speech may be attributed to the limited setting in which they have acquire the L2 (e.g. a classroom setting). Many of the aforementioned studies of DMs in L2 draw on methods from corpus linguistics to facilitate the quantitative and qualitative analysis of their data. In the next section, I discuss in more detail how corpus methods have contributed to the study of L2 pragmatic competence.

2.4. Corpus linguistics (CL)

2.4.1. Significance of corpus linguistics

A *corpus* (pl. corpora) is the main resource in corpus linguistics (CL). The term comes from the Latin word for "body". While the term can be used generally to refer to any body of text, in CL specifically, a corpus is "a collection of sampled texts, written or spoken, in machine-readable form which may be annotated with various forms of linguistic information" (McEnery et al., 2006: 4). Corpus linguists employ computers to analyse large sets of written or spoken data to answer questions in various areas in linguistics, including language teaching and learning (O'Keeffe et al., 2007; Aijmer, 2009; Granger et al., 2015; Haselow, 2021), discourse analysis (Biber et al., 2007; Baker & McEnery, 2015; Baker, 2023) and pragmatics (Adolphs, 2008; Aijmer & Rühlemann, 2015; Romero-Trillo, 2018). The application of corpus methodologies in different areas of linguistics (e.g. syntax,

semantics or pragmatics) has given rise to new sub-branches of the field (namely, corpus-based syntax, corpus-based semantics and corpus pragmatics; McEnery & Wilson, 2001:2). These methodologies have become beneficial tools for examining a range of linguistic patterns, including pragmatic features such as DMs.

CL is variously known as an *approach* (e.g. McEnery & Hardie, 2012; Brezina & McEnery, 2020), a *methodology* (e.g. Svartvik, 1992; Leech, 1992; McEnery et al., 2006) and a "quantitative paradigm grounded in the empirical tradition of language analysis" (McEnery et al., 2019: 74). To some corpus linguists, a corpus is more than a research tool and CL is considered to be a separate discipline that has a "theoretical status" (Tognini- Bonelli, 2001:1). Tognini- Bonelli (2001) proposes two different terms to differentiate two different views of CL: the *corpus-based* and *corpus-driven* approaches. In a corpus-based approach, a corpus is used to test and exemplify theories that have already been formulated. Therefore, the role of the corpus is mainly to provide evidence that can add minor modifications to existing theoretical models (Tognini- Bonelli, 2001; 65). On the other hand, in corpus-driven approaches, the corpus is the main source for finding evidence and developing new theories. Alternatively, many linguists believe that CL is mainly a method to investigate empirical data (see McEnery & Gabrielatos, 2006; McEnery & Hardie, 2012). McEnery & Gabrielatos (2006:37) argue that since CL is ultimately an approach to study language, it is impossible to have an "atheoretical approach". This means that CL is "a means to an end rather than an end in itself" (McCarthy & O'Keeffe, 2010:7). Following these linguists, I opt to define CL as a methodology for studying language by analysing authentic language examples on a large scale.

From this perspective, CL is a collection of analytical procedures that enable us to explore linguistic features in depth, through empirical investigation rather than relying on intuition (McEnery et al., 2006: 6; Hardie, 2016: 503). Corpora are sources of evidence which can be investigated to support or challenge linguistic theories. This empirical basis increases

reliability and validity in corpus research. In addition, CL benefits from the capacity to interrogate large collections of empirical data; this enables linguists to provide strong evidence, adjust existing theories and discover new theories in language use (Svartvik, 1992: 9; McEnery & Hardie, 2012: 1). Corpora, then, contribute to the advancement of various language theories "which were at best difficult to explore prior to the development of corpora of suitable size and machines of sufficient power" (McEnery & Hardie, 2012: 1).

The field of CL gained popularity in the 1980s, when computers for research became widely available to linguists (Hardie, 2016: 503)[4]. Computerised corpora provide researchers with realistic quantitative information about frequencies of linguistic phenomenon, which has, in turn, helped linguists in counting and observing new categories in language use (Hunston, 2002: 2). In comparison with manual analysis, computers "ensure consistency and minimise the impact of human error and subjective bias" in linguistic analysis (Collins, 2019: 2). The expansion of the field of CL has also allowed researchers to examine new linguistic models and draw on advanced research methodologies (Aijmer & Altenberg, 1991: 3).

The advancement of spoken corpora in the field of CL makes it possible for researchers to store and re-use large masses of spoken data that are collected or transcribed by others (Leech et al., 1995: 6). It also assists researchers in testing "the possible relationship between lexico-grammar, utterance function and discourse context, and to explore possible patterns in this relationship" (Adolphs, 2008: 1). This is beneficial for many linguistic fields such as pragmatics, discourse and L2 studies as researchers can compare their own transcribed spoken data with available spoken corpora. The capabilities of CL tools in the study of pragmatics, for instance, can be further enhanced by the inclusion of relevant metadata and pragmatic annotation that provides a record of the interactional context (such as

[4] The field of CL began in the 1950s, the era of post-Bloomfieldian structural linguistics in the United State of America. The field continued to expand despite criticisms by Noam Chomsky and advocates of his theories. (For details on the diachronic development of CL see Leech, 1991; Kennedy, 1998; McCarthy & O'Keeffe, 2010; among many others).

paralinguistic features) that may have been significant in signalling and interpreting the communicative purpose.

CL is particularly important for pragmatic studies that rely on looking at authentic language in context (Aijmer & Rühlemann, 2015: 12). CL brings together the study of macro-level characteristics through quantitative analysis and micro-level features through qualitative analysis (Reppen & Simpson-Vlach, 2020: 93). For instance, in the context of studying pragmatic features such as DMs, corpus tools can provide statistical information about preferences for use of some particular DMs over others. Qualitative analysis can be used to uncover the various functions performed by a single DM in different contexts, and to interpret how interlocutors employ DMs to evaluate their arguments (Schiffrin, 1987: 67; see section 4.3). Thus, CL has become one of the approaches utilised in pragmatics to examine the forms and functions of many pragmatic features (Andersen, 2011: 588).

The relationship between language form and utterance function is not one-to-one as assumed in pragmatic theories (Adolphs, 2008: 6). In other words, one lexical form can carry multiple functions (as is the case for DMs, see section 2.3.4) and equally, one communicative function can be performed through different forms. In the field of CL, most of the pragmatic studies are form-based, and they examine words or constructions according to various functions (Aijmer, 2018: 555). Researchers can identify all the forms in the data through employing corpus methods (O'Keeffe, 2018: 587). When studying DMs specifically, corpora can be used to identify all instances of DMs, their contexts and their positions. The researcher can then manually interpret their communicative functions by reference to the co-text. Although employing function to form approaches in corpus studies is still developing (O'Keeffe, 2018), there are many promising studies in analysing speech acts through CL (e.g. Adolphs, 2008; see also Collins, 2019: 13 for more examples of such studies).

In L2 studies, corpora are compiled to function as sources of observational data to "provide evidence about language use by both L1 and L2 speakers" (McEnery et al., 2019: 74). Researchers in L2 pragmatics can use corpus tools to determine the context of use and the functions of whatever pragmatic features are under consideration (Fernández & Staples, 2020: 243). Furthermore, reference corpora of L1 speakers can be used by researchers to compare examples of language use in their own corpora with native-speaker usage. Examples of such reference corpora include: the Cambridge and Nottingham Corpus of Discourse (CANCODE); the original 1994 British National Corpus (BNC) (Aston & Burnard, 1998); and the Spoken BNC2014 (Love et al., 2017).

CL offers many tools that are beneficial in examining child linguistic development (Demuth, 2008: 199). The advancement of CL in child language acquisition began to flourish in 1983, when Catherine Snow and Brian MacWhinney discussed the idea of constructing an archive for child language that would allow researchers to share their transcripts. Before the development of CL in child language acquisition research, parental diary-taking was the predominant systematic means by which data was sampled. (Behrens, 2008: XI). Corpus studies in child L1 and L2 acquisition were limited. One reason for this is the dominance at that time of the single case study approach, which focuses on one learner or a few learners. Another is that the field of language acquisition then focused on looking at limited linguistic features or only one register such as narrative (Biber et al., 1998: 173). Finally, longitudinal studies dominated in the field of child SLA (McEnery & Wilson, 2001: 3). These types of studies usually involve a significant amount of time observing a small number of children. Although the results from such studies are valuable, they often ignore the effect of individual variation in language development (Demuth, 2008: 199).

CL methods allow researchers to examine language use across multiple cases. Corpus tools permit analysis of large sets of data; this advantage of corpus tools provides a

foundation for the generalisation of results in child language studies (Biber et al., 1998: 173). Researchers can also benefit from using corpus tools to observe associations between linguistic features; thus, these approaches "enable a comprehensive description of language use at different developmental stages" (Biber et al., 1998: 173). Corpora of L2 data can be enhanced with information about the learners and/or the tasks. These details are valuable for researchers. They make possible the definition of subcorpora to compare groups of learners defined according to factors such as age, length of exposure, level of fluency or L1(Granger, 2002: 10). Thus, corpora of learner language must be designed carefully to capture and preserve this background data on the learners.

2.4.2. Key concepts in corpus compilation

2.4.2.1. Types of corpora

Corpus researchers distinguish types of corpora according to their purpose. There are two broad types of corpora, namely, *general* and *specialised* corpora. General corpora aim to broadly represent a language or language domain and are often known as reference corpora (as proposed by Sinclair, 1991: 25). This type of corpus is designed to be balanced and include samples representing a variety of registers. An example of this type is the British National Corpus (BNC), which comprises 100 million words. Another example is the Corpus of Contemporary American English (COCA, Davis 2009), which contains one billion words and continues to expand by 25 million words each year. The corpus includes eight different genres: fiction, spoken, magazines, newspapers, blogs, academic texts and other web pages. By contrast, specialised corpora aim to represent a sub-type of a language rather than a whole language (Hunston, 2002: 14), and is typically designed to answer questions about aspects of language that could be not addressed using general corpora (Reppen & Simpson-Vlach, 2020).

Both general and specialised corpora may be used in researching language teaching and learning. SLA researchers make use of a recent type of corpus known as *learner corpora*. A learner corpus is a type of corpora which consists of speech/writing by non-native speakers in their L2. Learner corpora are compiled to enable research on language acquisition, teaching and learning (Granger, 2002: 4). In building such a corpus, researchers will collect information on the differences or similarities among learners themselves or in comparison with L1 speakers. They may also enhance the corpus with information on the task the learners performed in producing the data (e.g. tests or interviews) and on the type of interaction (e.g. spontaneous or formal). Learner corpora can be large in size. An example of this type is the Cambridge Learner Corpus (CLC, Nicholls, 2003), which contains over 20 million words of anonymised scripts from the Cambridge language test, produced by students from 150 different countries. The CLC contains information about each learner's age, nationality, level of English and L1. Learner corpora can also be smaller. For instance, the Chinese Learner English Corpus (Gui and Yang, 2001) comprises approximately 1 million words; it is a specialised corpus of English scripts written by Chinese learners ranging from middle school students to senior English majors. As a specialised corpus in Chinese learners, the Chinese Learner English Corpus can be used to examine the differences between learners with a similar L1 but with different level of English and various age groups.

Other corpora that are designed for general language research purposes can also be useful in language learning studies. These include *comparable* and *parallel* corpora. Comparable corpora have the same corpus design and consist of samples of two or more different languages, or varieties of the same language. The function of such corpora is to allow comparisons of the different languages or varieties (Tognini Bonelli, 2012: 20). Parallel corpora, on the other hand, consist of a collection of the same texts in two or more languages. This may be a set of original texts and translations of those texts into other languages, or a set

of texts produced in two or more languages at the same time (Hunston, 2002: 15). Translators or learners can use this kind of corpus to search for equivalent expressions in the other language (Hunston, 2002: 15).

While many researchers utilise existing corpora, many also find they need to build their own specialised corpus (McEnery et al., 2006: 71). For instance, to research a specific register or specific time period, a tailor-made corpus may be desirable (McEnery et al., 2006: 71). If researchers, for example, want to study language in a highly specific context (e.g. informal interactions in a specific language community) which is not represented by any generally available corpus, they need to build their own specialised corpus (Love et al., 2017: 324), as indeed I do in this study (see section 3.2).

2.4.2.2. Corpus design and annotation

Once a researcher decides to build their own corpus, they must carefully consider its design. Corpus design can affect the subsequent analysis and consequently affect the reliability of the results (Reppen & Simpson-Vlach, 2020: 95). According to Rayson (2015: 32), the process of building and using a corpus has three main stages, namely, the compilation of a corpus, the annotation stage and data retrieval. In the compilation stage, a researcher should decide upon the design of the corpus and what material is to be included within it. For instance, a research question specifically about spoken language would dictate that the corpus should consist of speech and not writing. A corpus compiler should also consider first the optimal size of the corpus (Kennedy, 1998: 66). Although a large set of results can be advantageous, it also can posit a challenge for researchers carrying out qualitative analysis since they have to work with a massive number of results. However, researchers in many cases can benefit from these large numbers of examples through minimising them using the randomisation filtering tools or by constructing reduced subcorpora from these large corpora (Granger, 2021: 4).

On the other hand, small-sized corpora have certain benefits in specific conditions. Smaller corpora can be sufficient in searching for high frequent grammatical constructions (Davis, 2019: 76) and can be particularly beneficial if the researcher is looking to manually annotate the data to support their analyses (McEnery et al., 2006: 33). The optimal size for a corpus is dependent on the research questions to be answered. While researchers will benefit from including as much data as they can in their corpus, they should not be preoccupied with the size at the cost of principled design (Biber, 1993: 243; see also Davis, 2019: 77). No matter how large, a corpus will always constitute a sample of the language and so researchers should consider *what* the corpus represents in terms of an area of language, which will incorporate *how much* (Leech, 1991: 11-13; Leech, 1992: 107). Thus, we should look at other issues in designing the corpus including its r*epresentativeness* and *sampling technique*.

A collection of texts is not defined as a corpus in the strictest sense unless it represents a specific language or a variety of language. This representativeness differentiates a corpus from a text archive (Leech, 1991: 11), where the latter is simply a mass of data that is collected according to the available sources and not designed to be representative of any variety of a language. For a corpus to be representative, it should include the full variety of a language. In most cases, researchers can only reasonably collect a sample of the language they wish to study, and not the whole. The type of sampling frame used to accomplish this will have implications for the representativeness of the corpus. The sampling frame includes the type of language in a corpus (McEnery & Hardie, 2012: 8). If we are trying to build a representative corpus of written language, for example, the corpus should contain a variety of written samples including magazines, newspapers, books, letters, reports and all other written media (Hardie, 2016: 505). Reppen & Simpson-Vlach (2020: 95) list three different ways to capture representativeness in a corpus. First, researchers can seek representativeness of different registers, topics or discourse modes. Another issue to consider in conceiving

representativeness is the intentional choice of participants' nationality, gender, age or class. For instance, in SLA research, a corpus might sample language from a group of L2 speakers spanning different ages or different proficiency levels. The last category to include in seeking representativeness is the choice of a sample to be built on reception or production.

Sampling is often linked with *balance.* Achieving balance in the design of a corpus means ensuring that the appropriate variety of texts included are appropriate to the proportions of text types or speakers of the language variety being studied (Collins, 2019: 31). The idea of achieving perfect representativeness is debatable in the field of CL. Hardie (2016: 505) argues that perfect representativeness in a corpus is unattainable since even a large corpus such as the BNC would not capture all the possible types of texts. Thus, researchers should seek *better* representativeness instead of perfection (Hardie, 2016: 505). Researchers often face decisions between "what would maximise the representativeness and what is possible in practice" (Love et al., 2017: 326). For instance, although errors in L2 is part of learner performance, learners whose language is being sampled for a learner corpus may produce even more language errors due to pressure of being recorded during data collection that eventually affect the representativeness of a corpus as a sample of their actual performance. Hence, researchers should consider compiling a corpus that primarily fits their research purposes.

Once a researcher decides upon their research questions, they would eventually determine the content of their corpus to consist of writing, speech or both types. This decision will affect the necessary corpus construction procedures. For instance, in a spoken corpus, transcribing the data is an essential part of corpus compilation (Leech, 1991: 11). In compiling a corpus, either written or spoken, researchers can enhance their corpora with three kinds of information which may assists them in investigating their data: *markup*, *metadata* and *annotation* (McEnery & Hardie, 2012: 29). Markup is the process of inserting codes into

the text itself (McEnery et al., 2006: 22). In spoken corpora, markup might include, for instance, codes marking the points at which speakers begin or end their turns (McEnery & Hardie, 2012: 29).

Metadata provides details about the text itself such as author, the date of writing or recording and the type of language of a corpus (McEnery & Hardie, 2012: 29). In constructing a corpus for language learning and teaching purposes, metadata can be information about learners that enhances the analysis of a corpus with information about learners' variations (Gablasova et al., 2017: 137). In learner corpora, metadata may originate from a matrix design which researchers build upon constructing their corpus. This matrix consists of all the necessary variables in learners such as age, nationality, language level, gender, type of school (e.g. local, private, or international), location and class size and in teachers (e.g. degree of education, level of experience and their native language) (O'Keeffe et al., 2007: 1). This type of matrix works in collecting data from L1 learners. In researching L2 learners, Granger (2002: 9) illustrates that learner corpora are often compiled using a strict matrix design that includes more variables than these specified for L1 learners. The variables in L2 matrix design are specific to learners (e.g. L1, learning context, level of proficiency) and to the task (e.g. time, use of reference tools, audience) (Granger, 2002: 9). Using a matrix design of this sort is beneficial for L2 researchers to build subcorpora according to predefined criteria. These subcorpora can serve research purposes that include confirming or disconfirming theories about SLA, researching issues related to language transfer, or contributing to the production of language teaching materials and methods (Granger, 2002: 10).

Annotation is the second stage in corpus compilation after the process of data collection (Rayson, 2015: 38). Annotation is the process of identifying linguistic features by adding tags and codes. Examples of types of linguistic annotation include phonetic/phonemic

annotation, part of speech (POS) annotation, syntactic annotation (parsing), semantic and pragmatic levels of annotation. Leech (1997: 4-5) specifies two main advantages for using annotation in corpus analysis:

- Annotation facilitates the process of data retrieval. For instance, in an annotated corpus, it is easier to differentiate the multiple grammatically distinct elements (e.g. present tense, past tense, past participle, infinitive, noun and adjective) of a single form such as *set* through POS tagging.
- Annotations can also allow for the re-use of a corpus in subsequent research endeavours. The *re-usability* of corpora (as referred by Leech, 1997) makes a corpus available for various linguistic investigations other than the intended purpose for compiling it in the first place, by its creator(s) or other researchers (in case a corpus is made available to other researchers).

However, some linguists prefer the use of raw (unannotated) corpora (e.g. notably John Sinclair; cited in Leech, 2005: 26). Their argument is based mainly on two claims. The first claim is the interference of discourse with annotation codes, resulting in preventing the flow of discourse (Leech, 2005: 26). Second, such researchers may view annotation tools as "inaccurate or inconsistent" (Sinclair, 1991 cited in McEnery & Hardie, 2012: 14). The first claim can be refuted by using software programs that can hide the annotated form of texts and display the plain texts or raw data (McEnery & Hardie, 2012: 14). The other claim is countered through establishing standardised annotation tools that help in ensuring quality and consistency (McEnery & Hardie, 2012: 14).

2.4.4.3. Corpus Analysis

The four most basic corpus analyses have been identified as frequency analysis, keywords, collocation and concordance (McEnery & Hardie, 2012: 2; Brezina & McEnery,

2020: 16). Understanding the benefits of frequency lists requires an understanding of word *tokens* versus word *types*. A type is a particular word form considered in the abstract, whereas a token is a single concrete occurrence of some particular word form at a specific point in the corpus. Thus, a type is represented by many tokens (Hardie, 2016: 510). Word frequency lists present all the types in a corpus along with the number of occurrences (tokens) of each, typically ranked by frequency, highest first.

This frequency information can, for instance, be used in further analysis, such as comparing the frequencies of some set of items with the frequencies of the same language features in other corpora (in particular, large reference corpora). Putting corpus frequencies in such a comparative frame is valuable in SLA because frequency data normally cannot enhance our understanding of language development and use in L2 speakers without any context for how often we might expect a particular language feature to appear in general (Gablasova et al., 2017: 131). Corpora in general and in SLA field in particular are often have different sizes. Hence, researchers use relative frequencies/normalised values to report numerical data from these corpora (Gries, 2015: 52). These relative frequencies are valuable in case of comparing corpora of different sizes. McEnery & Hardie (2012: 49-50) illustrate that in order to determine the normalised frequency (*nf*) of a word in a corpus, researchers use a *base of normalisation* to perform the following calculation.

> *nf* = (number of examples of the word in the whole corpus ÷ size of corpus) × (base of normalisation)

This calculation will answer the question "how often might we assume we will see the word per X words of running text?" (McEnery & Hardie, 2012: 49-50). For instance, if we are looking at how many times the word type *like* occurs on average in each ten thousand words of the Spoken BNC2014, we must normalise its frequency to a base of 10,000 words. Originally, the word *like* appears 157,425 times in the Spoken BNC2014, a corpus which

consists of 11,422,617 words in total. The calculation for normalising the number of *like* occurrences is as follows:

$$nf = (157{,}425 \div 11{,}422{,}617) \times 10{,}000$$

The relative frequency of the type *like* is 137.8 per ten thousand words. Analysts should opt for a base of normalisation appropriate to the size of the corpora under investigation. Choosing the appropriate base of normalisation assists in preventing any misinterpretation of data and misleading the readers (Brezina, 2018: 43). For example, if a researcher is investigating a relatively rare language feature, say one that may appear only once a whole corpus, and chooses a normalisation base (e.g.1,000 words) higher than the total size of the corpus (e.g. 900 words), the relative frequency of this feature will be magnified (from 0.1 per 100 words to 1.1 per 1000 words). (Biber et al., 1998: 264).

Keywords are used to explore salience in a corpus or in Scott's (1997: 236) precise wording "unusual frequency" of words in a corpus. They can be used to compare items that are more or less frequent in two corpora, the target corpus (the data under investigation) and a reference corpus (Brezina & McEnery, 2020: 16). Keyword analysis looks for the words with unusually high frequency, i.e. *keyness*, in a corpus (O'Keeffe, 2007: 12). The top items on a frequency list, in most cases, tend to be grammatical words such as *the*, *and*, *is*, and *are*. In contrast, the keyword list is ranked according to how big the difference in frequency between the two corpora is. That means the top items are not the most frequent, but the most different. Analysis of keywords in L1 corpus data has been used by language teachers in constructing wordlists for teaching vocabulary in English for Specific Purposes programmes (O'Keeffe et al., 2007: 12). Keyword analysis in a learner corpus, meanwhile, makes possible the identification of words and phrases that are underused or overused by learners (Granger, 2009: 26), which can direct our attention to areas of development for learners.

Collocation is the co-occurrence of words in a corpus (Collins, 2019: 57) and represents another standard corpus analysis technique that can tell us about language use. Collocation can be observed informally, however, corpus analysis tools offer reliable statistical measures of collocation (Hunston, 2002: 68). Identifying collocation in a corpus is valuable for language teaching and learning purposes, because collocation plays a significant role in disambiguating the meanings of a word, and also because we can identify near-synonymous words through collocation (Xiao, 2015: 115). For instance, the word *tea* collocates with the modifier *strong* rather than *powerful* although the two words have similar denotational meanings (Xiao, 2015: 117). L1 speakers may experience difficulties in explaining the reasons beyond these co-occurrences of words. Thus, identifying collocation can assist language teachers with empirical evidence about when two verbs (e.g., *turn* and *go*) co-occur with certain preferences. For instance, things *turn* white or black but people *go* (not *turn*) mad (O'Keeffe et al., 2007: 59). A learner with a good knowledge of collocation can show a high level of proficiency in L2.

Concordancing is a central tool in corpus analysis (Rayson, 2015: 40). Concordances, also known as *Key Word In Context* (KWIC) (Aijmer & Rühlemann, 2015: 7), assist researchers in looking for all the different occurrences of a word in a corpus. In a concordance line, the searched for a word or phrase (node) is displayed in the centre of the computer screen, with the words that come before and after to the left and right. A special software program known as a *concordancer* is used to search the corpus and display the concordance. This software can often support other types of analyses as well. Some concordancers run locally on the analyst's computer, such as *WordSmith* (Scott, 1996) and *AntConc* (Anthony, 2023). Others work as 'a client server system over the internet' in which a client accesses a remote computer via a web browser, such as *SketchEngine* and *CQPweb*

(Hardie, 2012). This distinction applies to the tools generally, which go beyond concordancing.

Concordances enable in-depth qualitative analysis and support various sorting functions. They also help researchers to identify patterns in the use of the search terms (Baker & McEnery, 2015: 3). Analysts must carefully examine and interpret the examples within the concordance according to their theoretical framework and research needs in order to derive interpretations of the data. In the field of L2 teaching, concordances can be useful to teachers for building teaching materials to use in teaching tasks such as the search for idiomatic uses of words or teaching grammar inductively (O'Keeffe et al., 2007: 10). Indeed, many published grammar books make use of concordance lines to illustrate different uses of grammatical patters (see Biber et al., 1999; Carter & McCarthy, 2006; among others). These practices exemplify how the concordance is advantageous for researchers in the field of SLA as well as language teachers. In the present study, analysis of concordances will support me in examining the functions of DMs in the speech of the children under investigation.

2.5. Summary

In this chapter, I surveyed the literature on child SLA, DMs and corpus linguistics. I discussed the main factors that a researcher should take into consideration when researching child SLA in section 2.2. These include external factors (such as socio-cultural environment, educational setting, and duration of exposure to L2 input) as well as internal factors (such as cognitive maturation, age, effect of L1, and motivation). Researchers of child SLA should account for the effect of such factors on development of children's pragmatic competence in their L2.

In section 2.3, I presented different theoretical views on studying DMs, as well as the different approaches to labels, definitions, characteristics and functions of DMs that derive from these various perspectives. I have established that while there has been extensive

research into DM acquisition by adult L2 learners, there is a scarcity of studies investigating frequencies and functions of DMs in the speech of L2 children. The study of DMs in the present thesis focuses mainly on the speech of Arab children learning English as a L2. Moreover, the study is driven by exploration of the various DM functions, as outlined in the textual and interpersonal domains of functions suggested in previous studies.

In section 2.4, I presented an overview of CL, the principal methodology of the present study. This overview introduced some basic terms in the field and highlighted the significance of corpus approaches in the field of SLA and pragmatic competence. In particular, I looked at procedures for building and analysing a corpus in section 2.4.4. In the following chapter, I will discuss the compilation of the corpus of the English L2 speech of Arabic-speaking children that is the main data of the study, as well as a reference corpus of speech by L1 English-speaking children.

Chapter 3: Methodology

3.1. Overview

In this chapter, I explain decisions made in regard to the methodology of preparing, collecting and analysing the data in the present study. In section 3.2, I discuss the compilation of a corpus of L2 English speech by Arabic-speaking children, which I name the *Ar-Children* corpus. I also describe the steps by which a subcorpus of child language from the Spoken BNC2014 was constructed, in section 3.3. Then, I describe the compilation of various subcorpora from the Ar-Children corpus that will be used to account for differences among the speakers, especially the time they spent studying preschool in English-speaking countries. In section 3.5, I outline the procedures through which I identified the selected DMs to investigate. I also provide a detailed account of the framework I applied to document and analyse the functions of DMs.

3.2. Construction of the Ar-Children corpus

To investigate the frequency and functions of DMs in the speech of Arabic-speaking children, I compiled a corpus consisting of authentic peer talk in English. Corpora of the language of adult learners of English are abundant, but corpora of child learner English are much less so. To my knowledge, no corpus specifically documenting the language of Arabic-speaking children learning an additional language is available. Some Arab child L2 corpora do exist, but their focus is the acquisition of a single or a limited number of speakers; these corpora are designed mainly to study language development by observing children for a lengthy period of time (i.e. longitudinal studies). As such, these corpora have been constructed using elicitation methods (e.g. story retelling or picture description) to capture language data, rather than recording natural usage. An example of such corpus examining a

single Arab child is Alhindawy (2016). Since no corpus consisting of natural data sampled from a number of Arab children learning English as a L2 exists, I set about compiling my own corpus. In this section, I explain in detail the process of compiling this corpus of the language of Arabic-speaking children (hereafter Ar-Children corpus), from recruiting participants to collecting, transcribing and coding the data.

3.2.1. Participants

3.2.1.1. Participant background

Education in the Kingdom of Saudi Arabia is provided through public, private or international schools. In the public sector, English teaching is limited to only two sessions per week. The private and international sectors, on the other hand, prioritise English teaching in their schools, but to different degrees. While the private sector offers only five English classes per week, in international schools (also called *English Medium Instruction* schools), the whole curriculum is presented exclusively in English (either British or American). In these international schools, use of the children's L1 Arabic is limited to one or two sessions per day. Moreover, children are discouraged from using their L1 in communication inside or outside the classroom during the school day. The schools from which the data in my corpus was compiled are all American international schools, of which there exist a greater number in Saudi Arabia than of British international schools.

3.2.1.2. Participant recruitment

Prior to data collection, I contacted a number of international schools in Riyadh, the capital of Saudi Arabia, and got permission from three schools to attend some of their classes and recruit participants for the study. From these three schools, I recruited 120 female school-age children, according to criteria detailed below. The number of participants was dependent

on the number of the girls whose families provided approval for their participation. The participants come from various dialect backgrounds, as indicated by the variety of Arabic nationalities represented in the corpus, and illustrated in table 3.1. Most notably, three quarters of the participants are Saudi or Egyptian.

Table 3.1. *Number of participants in the Ar-Children corpus according to nationality*

Nationality	**No. of participants**
Saudi	65
Egyptian	24
Jordanian	9
Syrian	7
Sudanese	4
Palestinian	3
Yemeni	3
Lebanese	2
Tunisian	2
Kuwaiti	1
Total	**120**

The participants in the study are all girls. The decision to recruit only female speakers was motivated by three factors. First, many studies have shown that DMs are more frequent in female speech than male (e.g. Escalera, 2009; Odato, 2013; Bu, 2013) and so this would maximise the opportunities for identifying and analysing the various forms and functions of DMs. Second, Saudi Arabia's education policy does not allow mixed-gender classes; it was

therefore more practical to collect data from speakers of one gender only. Finally, in recruiting participants of a single gender, I removed gender as a potentially confounding variable in the analysis.

The main objective of the present study was to collect data from school-age children. This meant that the sample of school-age children could range from first graders – who are often between 5 to 6 years old – to year six children, whose age typically ranges from 11 to 12 years old. I decided to include only juniors, whose age ranges between 9 to 12 years old. This choice was motivated by the following factors. First, previous research suggests that children at this particular stage move from being parent-centred to being peer-centred (Labov, 2001: 101; see also Kerswill, 1996:196, and Kerswill et al., 2013: 263) in terms of their language usage. The importance of this is that peer talk is associated with a greater diversity of DM usages (see Levey, 2006). As discussed earlier, studies focused on children should take into account external factors that may affect their language acquisition including contextual factors such as peer talk (see section 2.2.2.2).

Additionally, for the purpose of comparative statistical analysis between L1 and L2 speakers, I opted to limit the study to a single age group. Focusing on recruiting a larger number of participants in the same age range offers a better-focused comparison rather than sampling smaller groups from different age ranges. Age, as an internal factor, can influence child language acquisition (see section 2.2.4). Although this factor was not a variable of direct interest in the present study, I sought to control its effect by including participants from only one age group.

3.2.2. Data collection

3.2.2.1. Pre-collection phase

As this study involves human participants and, specifically, minors, I obtained ethical approval to conduct the research from Lancaster's University Research Ethics Committee (UREC) prior to data collection. Then, with headteachers' permission, children and their families were asked to participate in the research through an invitation notice and participation flyer (Appendix A). Children willing to participate, who had their family's approval, were given information sheets (Appendix B) and a background questionnaire (Appendix C) for their parents to complete. These papers were also provided in electronic form, and were available in both English and Arabic.

Prior to the actual data collection phase and with consent from the headteachers, I informally observed some online speaking classes at one of the schools. These are extracurricular sessions held twice a week to promote speaking skills. In these sessions, teachers usually assign a specific topic of conversation. The children are then required to develop their own ideas in either paired or group conversation. I observed that these sessions were purposely unplanned and children were encouraged to ask their own questions to their partners without any conversational prompts. These observations were valuable in determining the best way to collect data from the participants, and estimating how much data could be collected. The main observation from these sessions was that children were able to carry on conversations without any guidance if matched with their close friends. It was apparent, however, that some of the children nevertheless benefitted from guidance from their teacher. Hence, I decided to design prompt cards to help those who might need such guidance during the actual data collection. These prompt cards were derived from conversational topics covered in the sessions I had observed (and are given in Appendix D).

The practical restrictions of time and the fact that I was a single researcher placed limitations of how much data I could collect. Nevertheless, I endeavoured to collect approximately 100,000 words, which I considered sufficient to offer a picture of children peer talk in English-medium schools. To construct a corpus of this size, I had to estimate the number of words in each text sample to determine the optimal size of participants' contributions. According to Smith (2019: 34) the estimated number of transcribed words in one hour of recorded conversation is 8,000 words. This means that speakers produce at least 100 words per one minute. Taking this estimation into account, I planned to record each pair of children for at least ten minutes to get a sample that would result in transcribed conversations of approximately 1000 typed words from each pair. The estimation of text sample is motivated by the fact that DMs are frequent linguistic features of spoken discourse. Leech (2007: 141) illustrates that a text with a word count of 1,000 is sufficient for frequent grammatical structure.

As previously mentioned, parents were asked to complete a child background questionnaire prior to any data being collected. One of the objectives of the present study is to examine the role of the context of preschools and length of time spent in these preschools on the acquisition of DMs. For this reason, I requested that families answer some questions about their children's experiences in preschools in the background questionnaire. The questionnaire revealed that over 92% of participants had previously been enrolled in one of a number of different types of preschools. Only eight girls had not attended any type of preschool. Attendance at these preschools took place in a range of countries, including Arab countries (mainly Saudi Arabia), and English-speaking countries including the United States of America, the United Kingdom and Australia. Figure 3.1 shows how different contexts of preschool attendance were categorised in the analysis of questionnaire responses.

I considered only two variables for categorising preschool experience. The first variable was the language of the country in which the preschool experience took place, illustrated in the centre column in figure 3.1. The other variable was the length of time spent in preschool in an English-speaking country, illustrated in the three green boxes right to the "English-speaking countries" box in figure 3.1. These are the only variables relevant to the research questions at hand. The length of time in preschool in Arabic context is not a variable which requires analysis. This is because the majority of participants are from Saudi Arabia (see section 3.2.1.2) in which formal enrolment of preschool is only one year. Information about the length of enrolment for the other children who studied in other Arabic-speaking countries would not yield valid results due to their small number compared to the Saudi children group. In addition, the types of preschool within the broader contexts considered (e.g. within Arabic-speaking countries) were not taken into account. These types are presented in the first three boxes on the right column in figure 3.1. The first grey box "Arabic only" presents the type of preschool in which children only Arabic curriculum is being taught. This type is excluded from the analysis of preschool effect on English DM acquisition due to their lack of English knowledge prior to join the primary school (see section 3.4). The second and third orange boxes in figure 3.1 are "English only" and "English and Arabic" are treated as one group. These two types focus on English as a medium of interaction with different amount of exposure. I treated these two groups as one because there is no information about the amount of exposure to English in each single participant's experience. This means that even if they were taught in English in both types, Arabic is still present as the language of the country and spoken by teachers, school staff and other students.

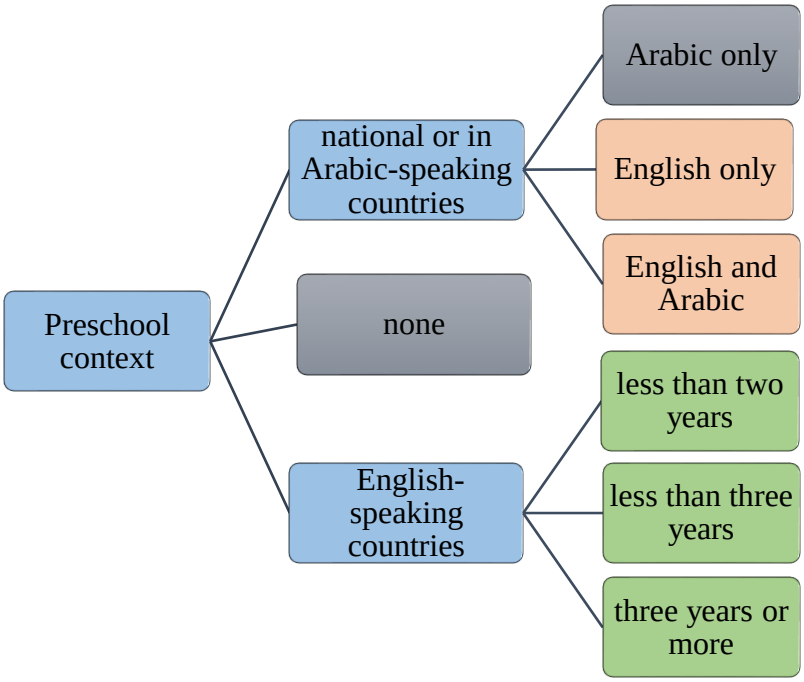


Figure 3.1. *Categorisation of participant children's preschool experience*.

3.2.2.2. Data collection phase

I devised a set of criteria by which to determine whether a child participant's speech should be included in the corpus or not. Only children who met the criteria for inclusion were included in the study. The criteria for participant inclusion were as follows:

1. The participant was a native speaker of Arabic. This excluded children who spoke a language other than Arabic at home. However, I counted children who spoke two or more languages at home, one of which was Arabic, as native speakers for the purpose of this study.
2. The participant had studied at an English-medium school from the first grade to the current time. This excluded children who attended either of the other types of school (national or private sector) earlier in their education. This will ensure the homogeneity of the participants and maximise the reliability of the results.

The data were audio-recorded in a quiet room on school premises. Recording sessions were done mostly during the extra-curricular sessions (these sessions are free classes for children to build their social and creativity skills) and/ or sometimes during the speaking sessions (these classes are part of the English curriculum). I recorded data from all participants as they engaged in peer talk. Each participant was paired with another participant who was a close friend and both were invited to sit at a table filled with items (e.g. books, paper, pens, and an electronic tablet) along with the recorder. The prompt cards were randomly organised in a box located at the centre of the table. I spent five to ten minutes with each pair to explain the process of recording, answer their questions and try to help them become settled in the situational context prior to the actual recording. My being present in the room allowed me to take notes about the participants that eventually would assist me in recognising their voices in the transcription phase. I planned to record each pair of participants for ten minutes (see section 3.2.2.1). However, not all the recordings ended up being this long. Some pairs appeared to be inhibited by the presence of the recorder, although most were not. Less than half of the participants (n=50) used the prompt card box, with the majority (n= 70) choosing to select their own conversational topics. This makes the data a mix of unguided and semi-guided peer conversations. The recordings of peer conversations resulted in the Ar-Children dataset of 60 samples of conversations, which when transcribed produced 78,035 words.

3.2.3 Transcribing the data

3.2.3.1 Phase one: orthographic transcription

All sixty recorded conversations were transcribed orthographically, and the transcriptions were checked manually for errors or missing words. The text files were labelled according to the time and date of the recording (e.g. file 1103 recorded at 11:03 AM

on February 6th). Each speaker was given a unique ID. Transcription conventions (Appendix E) were adopted from the Spoken BNC2014 manual (Love et al., 2017).

3.2.2.3. Phase two: marking utterance boundaries

One element that plays a crucial role in deciding the function of a DM is its position within the utterance or turn. As noted previously (section 2.3.3), DMs are flexible and may appear in any utterance position (Schourup, 1999; Brinton, 2017). Hence, it was necessary to mark up utterance boundaries in the transcripts. Drawing upon the framework of segmentation suggested by Degand & Simon (2009) and other literature on defining utterance boundaries (e.g. Othman, 2007; Jun & Fougeron, 2001), I established rules for utterance segmentation. These rules refer primarily to prosody (the intonation contour) and secondarily to syntax (phrase dependency).A stretch of speech needed to meet the requirements at either prosodic level, the syntactic level or both, to be marked as a complete utterance unit.

The rules for utterance segmentation were as follows:

1. At the prosodic level, an utterance is deemed to end when it reaches a point "marked by a major continuation rise or a major final fall" or "by a large final lengthening and is optionally followed by a pause" (Jun & Fougeron, 2001: 220).
2. At the syntactic level, an utterance is deemed complete at the point where all the governed complements or dependents of the clause's main verb are themselves complete (Degand & Simon, 2009: 7). However, the completion of a full independent clause cannot be the only syntactic criterion for a syntactically complete utterance, for two reasons. First, subordinate clauses are not governed by a main verb but nevertheless, may constitute a full prosodic unit in themselves. The other reason is that turns in spoken discourse often include incomplete clauses. Speakers may often use noun, adjective or prepositional phrases amid their speech that are not linked to

any main verb. I deem such extra-clausal phrases and subordinate clauses to be complete utterances if, considered alone, they meet the prosodic requirement (stated as rule 1).

3. DMs are considered to be part of the utterance unit to which they are adjacent, unless they are followed by a pause. In this case, the DM is deemed to be a single utterance by itself. This rule does not apply for a DM in turn-final position as it is counted as part of the utterance it terminates.

Utterance segmentation was accomplished in two phases. In the first phase, I used PRAAT (Boersma & Weenink, 2001) to analyse intonation contours, silent pauses and final lengthening, allowing me to mark utterance boundaries. In this phase, I annotated intonation contours according to the following conventions:

- A full stop (.) marked falling intonation contours at the end of utterances.
- A slash (/) marked rising intonation contours.
- A comma (,) marked a level intonation contours.
- Incomplete or truncated utterances or words were marked with a hyphen (-).

In the second phase, I reviewed this initial utterance boundary markup to take account of syntactic phrase boundaries (including noun phrases and preposition phrases) and clause boundaries. The corpus and accompanying speaker metadata were indexed in CQPweb (Hardie, 2012). The speaker metadata is drawn from the background questionnaire and includes nationality, age, context of studying preschools, number of years spent in preschool and language of preschool. This metadata allows for restricted CQPweb queries, such as limiting a search to within the speech of only specific speakers. It also makes it possible to compile subcorpora for use in exploring the effect of external factors (see section 4.3).

3.3. BNC2014 children subcorpus

The main purpose of this study is to find out how Arabic-speaking children use DMs in their L2 speech in comparison with the use of DMs by L1 speakers (see section 1.2). Performing such a comparison requires L1 speaker data. Ideally, the data would be a sample of the exact same type, i.e. a corpus collected in the same way, with children matching the Ar-Children participants for everything except L1/L2. However, this was not possible due to time constraints. The alternative option was to use data from one or more existing and available corpora of L1 children. The available L1 school-age children corpora that I found were constructed for specific purposes or unavailable to access online. For instance, the Lancaster Corpus of Children's Project Writing (LCCPW) is built to examine school-age children' knowledge on writing, so was the wrong specific type of child language for my purposes. The Polytechnic of Wales (POW) corpus consists of spoken conversations from 120 children with an age range of 6 years to 12 year old; thus, the sample of participants in the corpus is a good match for my data, in terms of speaker profile. However, I decided against using this corpus for two reasons. First, the data in the corpus was collected between the years 1978 and 1984. This makes it potentially unrepresentative of spoken modern-day children's English, so that using it might yield unreliable results; language is of course prone to change and children and teenagers are among the leaders in pioneering such changes (see section 2.2.2). The other reason is that the POW corpus is not available to access online. Most existing spoken corpora that *are* available online have been built with a primary focus on recording children under the age of five, in order to investigate their first language acquisition[5], and related issues such as language impairment, phonological acquisition, and literacy. What remained after eliminating these possibilities was the Spoken BNC2014,

[5] One primary repository of such corpora is CHILDES: see https://childes.talkbank.org/access/Eng-UK/.

which despite being a general spoken corpus, rather than a corpus of child speech particularly, offers the closest match to the Ar-Children corpus in the sense that among its speakers are numerous L1 English speakers of comparable age to the speakers in the data I had collected.

Although the Spoken BNC2014 was hence the optimal choice, in the sense of being the only workable option, the context of the child speech that it contains – being adult-led conversations – differs from the peer conversations in the Ar-Children corpus. Theoretically, adult-child interactions are often asymmetrical, featuring scaffolding by the adults and use of response markers from the children. By contrast, interaction between children is more symmetrical and likely to feature co-equal negotiation around, for example, turn-taking (see section 2.2.2). However, many studies on DMs have shown that the context of interaction has no effect on the frequency of DM use (see e.g. Bu, 2013; see further section 2.3.5). Although, then, I do acknowledge the differences in their structure and functions between adult-child and child-child interactions, I nevertheless opted to use the child speech from adult-child conversations in the Spoken BNC2014 to construct a subcorpus comparable to my own data because it provides the closest available source of age-matched child language use. But the differences between the situational contexts captured within the two corpora mean that direct frequency comparisons must be interpreted with caution.

The Spoken BNC2014 consists of 11.5 million words of transcribed spontaneous conversations among L1 English speakers (Love et al., 2017). It incorporates speaker metadata covering a wide range of factors, including age and educational background. I manually checked the metadata to select the IDs of speakers of the same ages as the participants in the Ar-Children corpus. Using these speaker IDs, I constructed a subcorpus in CQPweb (Hardie, 2012) that comprises six speakers (3 male, 3 female) and consists of

134,316 tokens. This collection of utterances is referred to henceforth as the *BNC2014 children subcorpus*.

3.4. Compilation of Ar-Children subcorpora

One of the objectives of the present study is to test the effect of certain external factors (context and length of studying preschool) on the acquisition of DMs by L2 school-age children. To assess these effects, I needed to separate out speakers according to these factors, so that the DMs I extracted using corpus queries could be classified according to the speaker groups. To do this, I separated the speakers into Arabic countries preschools and English countries preschools. Then, I excluded all participants who did not attend preschool or attended an Arabic medium preschool (N = 9). The use of DMs by each participant was calculated individually (see Appendix F). Then, I grouped these frequencies based on the speaker's experience in preschool, as specified below. At this point, I decided to exclude those who spent three years or more in preschool. Participants in this category were too few to be a representative sample (N = 4).

In addition, I tried to control for one of the internal factors that might affect the acquisition of L2, that is, age. It has been demonstrated in many studies that age has a significant effect on SLA (e.g. Colletta et al., 2015; Flores et al., 2017; Sopata & Dlugosz, 2022; Paradis, 2023). From these two steps together, I compared the frequencies of DMs according to the context of studying preschool, establishing the following groups:

a) Zero English-speaking experience, abbreviated as APZ: speakers in this group attended preschool inside Saudi Arabia or in any other Arabic-speaking countries (N = 60) but not in English-speaking countries.

b) One-year English-speaking experience (AP1): in this group, participants spent at least one year but less than two years of their preschool in English-speaking countries (N=26).

c) Two-year English-speaking experience (AP2): participants spent at least two years but less than three years of their preschool in English-speaking countries (N=21).

To perform the comparison, I compiled six different subcorpora from the Ar-Children corpus, each consisting of participants with similar preschool experience and age. These subcorpora were constructed through specifying the required type of participants based on their experience in preschools (see figure 3.1) and according to their age, i.e. 9-10 and 11-12 years old. The resulting subcorpora are presented in table 3.2.

Table 3.2. *The subcorpora compiled from the Ar-Children corpus*

Subcorpus	**No. of participants**	**Size in words**
9-10 Arabic countries preschool	38	17,538
11-12 Arabic countries preschool	22	12,982
9-10 one year in English-speaking countries	16	11,086
11-12 one year in English-speaking countries	10	7,820
9-10 two years in English-speaking countries	14	10,367
11-12 two years in English-speaking countries	7	4,506

I used these subcorpora to determine the number of clauses produced by each group of participants, as in table 3.3. The process for identifying and counting clauses is described in details in section 3.5.3.

Table 3.3. *Total number of clauses in the six subcorpora from the Ar-Children corpus*

Pre-school group	**Age: 9-10**	**Age: 11-12**
APZ	2,380	1,732
AP1	1,583	1,460
AP2	1,084	667

3.5. The procedure of analysis

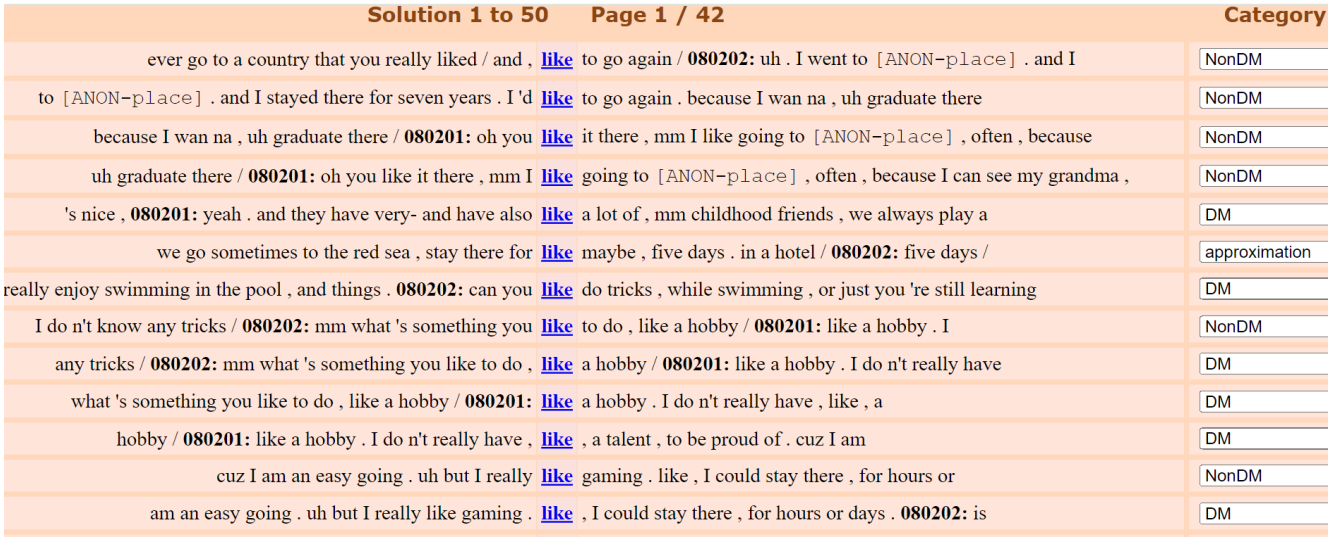
3.5.1. Selection of DMs for analysis

DMs are grammatically optional elements that originate from different syntactic word classes (see section 2.3.1). These expressions are flexible, since they have weak clause association and can occur in any utterance position (see section 2.3.3). Crible (2017: 108) attempts to identify a list of DMs in English according to these criteria; the resulting list of items includes most of the DMs dealt with by previous literature, such as *actually, although, and, anyway, as, you know, because, however, I mean, in fact, now, oh, okay, or, right, so, well, yeah*, among others. In the present study, I adopt Crible's (2017) list of DMs as the set of items of interest for my research questions.

To identify the most frequent DMs in both corpora, a frequency list for each corpus was obtained using CQPweb (Hardie, 2012) (see Appendix G). Then, I manually extracted all words able to function as DMs that were among the fifty most frequent words. The words in question are *and, like, yeah, so, okay, but, oh, because/cos* and *well*, I supplemented this list with two DMs that are formed by a combination of words, i.e. *you know* and *I mean*. The reason behind adding these two markers is that they are among the most studied DMs in the literature, though they would not appear on a frequency list generated according to single tokens. I then decided to exclude three words that can function as DMs from the analysis

despite their appearing with high frequencies in both corpora: the conjunctions *and*, *but* and *cos/because*. These are excluded because a high proportion of instances of these words were determined to be non-DMs, based on manual checking of the concordance lines.

The resulting list of DMs to be analysed consists of eight items: *like, yeah, okay, well, oh, so, I mean* and *you know*. The frequencies of these items, as DMs and with other functions, in the two corpora are presented in section 4.2 and 4.3. In each case, the differentiation of DM and non-DM uses was carried out using the categorise query function in CQPweb (Hardie, 2012). As illustrated in figure 3.2, this function allows each result in a concordance to be tagged manually with one of a list of user-defined categories. Tallies of the different categories can then be extracted, and the results separated out into concordance lines for each category for subsequent analyses.



Solution 1 to 50	Page 1 / 42	Category
ever go to a country that you really liked / and ,	like to go again / **080202:** uh . I went to [ANON-place] . and I	NonDM
to [ANON-place] . and I stayed there for seven years . I 'd	like to go again . because I wan na , uh graduate there	NonDM
because I wan na , uh graduate there / **080201:** oh you	like it there , mm I like going to [ANON-place] , often , because	NonDM
uh graduate there / **080201:** oh you like it there , mm I	like going to [ANON-place] , often , because I can see my grandma ,	NonDM
's nice , **080201:** yeah . and they have very- and have also	like a lot of , mm childhood friends , we always play a	DM
we go sometimes to the red sea , stay there for	like maybe , five days . in a hotel / **080202:** five days /	approximation
really enjoy swimming in the pool , and things . **080202:** can you	like do tricks , while swimming , or just you 're still learning	DM
I do n't know any tricks / **080202:** mm what 's something you	like to do , like a hobby / **080201:** like a hobby . I	NonDM
any tricks / **080202:** mm what 's something you like to do ,	like a hobby / **080201:** like a hobby . I do n't really have	DM
what 's something you like to do , like a hobby / **080201:**	like a hobby . I do n't really have , like , a	DM
hobby / **080201:** like a hobby . I do n't really have ,	like , a talent , to be proud of . cuz I am	DM
cuz I am an easy going . uh but I really	like gaming . like , I could stay there , for hours or	NonDM
am an easy going . uh but I really like gaming .	like , I could stay there , for hours or days . **080202:** is	DM

Figure 3.2. *Display of a CQPweb categorised query for* like *in the Ar-Children corpus.*

Common non-DM uses of the various items include their functioning as grammatical elements with a role in the syntactic clause. Table 3.4 offers examples for the eight frequent words in their occurrences as syntactic elements in the Ar-Children corpus and the BNC2014

children subcorpus. Beside their syntactic use, some other examples are coded as non-DMs in both corpora or in one of them. For instance, although *like* as a quotative and approximative marker has been studied as a DM (e.g. Meehan, 1991; Romaine & Lange, 1991; Tagliamonte & Hudson, 1999; Buchstaller, 2001; Andersen, 2001; Schweinberger, 2014; Tagliamonte, 2016; Beeching, 2016; D'Arcy, 2007, 2017), all instances of *like* as a quotative and approximative marker are listed as non-DM in the present study. This is because their absence will affect the grammaticality of the sentence, thus, failing to achieve one of the essential features of DMs (see section 2.3.3). With regard to *oh*, there are some examples of *oh* in the BNC2014 children subcorpus that are excluded from the analysis, in which *oh* is uttered by participants whilst playing games (to show excitement) or chanting songs. These instances of *oh* are coded as non-vocative examples of non-DM use.

Table 3.4. *Examples of non-DM uses of the eight DMs under study*.

DM	Non-DM function	Example (with source text ID / corpus)
like	verb	(1) so, do you ***like***/ the school. (0846/ Ar-Children)
	preposition	(2) that doesn't look ***like*** a lot of rice (SHP3/ BNC2014 children)
	conjunction	(3) and there in there and then you put their hand on their body ***like*** you're staying there (SX2V/ BNC2014 children)
	quotative	(4) I was ***like*** / what the heckjust happened (3929/Ar-Children).
	approximative	(5) we watched some episodes of Harry Potter over there. and ***like***, ten or eleven episodes (0918/ Ar-Children).
so	intensifier	(6) I found , it was , mm a slide , and swing , and that was , ***so so so*** amazing (1116, Ar-Children)

	fixed expression		(7) he invited other children to his party and ***so*** on (S5U8, BNC2014 children)
	conjunction		(8) they had them in they had a pair in the boot ***so*** that was a ten (SA2J, BNC2014 children)
	substitute to avoid repetition		(9) S0653: have you got still got a room full of bunny rabbits --ANONnameF? S0654: yeah S0655: ***so*** have I (S24D, BNC2014 children)
	verb substitution		(10) S0417: sodium chloride is n't it? S0416: I think ***so*** (SV5A, BNC2014 children)
yeah	response signal	yes/no question	(11) 085602: did you have your breakfast/ 085601: ***yeah.*** I did have breakfast (0856/ Ar-Children)
		tag question	(12) S0653: oh well officially I 'm a grown-up are n't I? S0655: ***yeah*** (S74A / BNC2014 children).
okay	adjective		(13) they're like. hello sweetie are you ***okay.*** (8350/ Ar-Children)
oh	fixed expression		(14) 122002: how much popcorn you ate, 122001: ***oh*** my goodness, no one, nothing, (1220/ Ar-Children).
	filled pause		(15) S0417: I 'm a gonna roll down the streets on --UNCLEARWORD S0418: >>mummy S0416: ***oh oh oh oh*** S0418: >>mummy S0416: >>***oh oh oh*** (S6KV/ BNC2014 children).
well	adjective		(16) and I went to a place called [ANONplace], it's / ***well*** known. I guess, if you know it. (1023/ Ar-Children)

	adverb	(17) I did n't really get some questions, but it was okay, like I did ***well***, I think (0946/ Ar-Children)
	fixed expression	(18) I 'm really excited as ***well***. (0848/ Ar-Children)
you know	declarative	(19) it is actually my favourite, like, ***you know*** the movie/ it is very nice/ (0902/Ar-Children)
	interrogative	(20) do ***you know*** what I was thinking? (S5XL/ BNC2014 children)
I mean	fixed expression	(21) you only realise it and then you fall down do you know what ***I mean*** ? (S3CK/ BNC2014 children)

In the Ar-Children corpus, it was possible to identify the DM or non-DM function of every instance of the eight items in question, because the audio recordings were available to consult in cases where the function was ambiguous in the transcription. But in the BNC2014 children subcorpus, some instances of each item had to be coded as unclear, since they were ambiguous in transcript and no audio was available. Instances were usually hard to classify if they occurred within incomplete turns or in the vicinity of unclear words that were left untranscribed, as exemplified in (22).

(22) S0653: only if you have a bit more potato

S0654: >>--UNCLEARWORD

S0655: ***okay*** I

S0652: >>potato?

S0653: >>and I would like you to stick (S47C / BNC2014 children)

3.5.2. Assignment of DM functions

After elimination of instances of the eight items with non-DM use (section 3.5.1), I downloaded concordances of the DM instances from each of the two corpora in CQPweb and inserted them into spreadsheets (using Microsoft Excel, as illustrated in figure 3.3). Within these spreadsheets, I assigned each concordance line a code representing its function as a DM. This step allowed subsequent calculation of the frequencies of different DM functions, as well as categorising the concordance lines by function for separate analysis.

7	80201	802	call us / me , I 'm , like ,	okay	, I 'm sorry , bye / uh , are	DM	quotative
8	80202	802	go back , now I 'll be in peaceful .	okay	, I mean / but of course I love my	DM	turn taking
9	80202	802	Ariana Grande , and I hate Billi Eilish oh ,	okay	/ mm I used to like Billi Eilish , I	DM	understanding
10	80202	802	the stuff that ariana Grande what 's he sings ,	okay	, uh do you use any socials. like / mm	DM	topic change
11	80202	802	gaming ones , like other than Roblox / nothing	okay	. mm okay . never mind do you wan na	DM	planning
12	80201	802	like other than Roblox / nothing . okay . mm	okay	. never mind do you wan na tell me about	DM	closure
13	80202	802	be proud . no-no six- forty five oh wow .	okay	. yeah six , there 's barely no time ,	DM	confirmation
14	82101	821	thank you , thank you . welcome , complete /	okay	, uh let 's talk now about another topic ,	DM	topic change
15	82401	824	na be , I gon na be a doctor ,	okay	, actually me I do n't know yet , there	DM	turn taking
16	82401	824	like that colour , See , purple with white ,	okay	, well . actually . what is your favourite thing	DM	floor holding
17	82401	824	are Ø¬Ø¯Ø§Ø± . and that , like that his Ù,Ø²Ø§	okay	. maybe he / for that , just a minute	DM	seek confirmati
18	82401	824	maybe the cat , uh / just a-just a minute	okay	. let me complete , okay . you know ,	DM	confirmation
19	82401	824	just a-just a minute okay . let me complete ,	okay	. you know , and after if you see it	DM	confirmation
20	82402	824	you really do n't like that , so complete .	okay	, I get it , that 's it , like	DM	closure
21	82401	824	, uh mm your sisters , or your brother .	okay	, I have I 'm gon na set the biggest	DM	turn taking

Figure 3.3. *Screenshot of a spreadsheet of a concordance of* okay *as a DM, annotated for different DM functions*

3.5.2.1. Rationale behind the framework of DM functions

The framework that underlies my assignment of codes for the functions of DMs, as described above, was constructed on the basis of a number of previous studies on DM functions, specifically Brinton(1996, 2017), Aijmer (2002) and Müller (2005). I have briefly discussed these scholars' frameworks previously in section 2.3.4. In detail, the systems of functions that they present are as follows:

- Brinton's (1996, 2017) classification:

Brinton (1996, 2017) classifies DM functions into three domains: *ideational, textual* and *interpersonal*. The *ideational* domain refers to the original use of the forms of DMs prior to their pragmatic functions; this is alternatively known as the "propositional" or "conceptual" use of DMs. For instance, one of the ideational function of *like* is a preposition meaning 'similar to'. This domain is referred to as the non-DM use in my study (see table 3.4).

The *textual* domain in Brinton's classification is the use of DMs to create coherence in discourse. Brinton lists six functions of DMs within the textual domain; these include "initiating or closing discourse, marking a boundary in discourse (new topic/shift in topic/resumption of topic), denoting new or old information, holding the floor or sustaining discourse, acquiring or relinquishing the floor, or marking sequential dependence" (Brinton, 2017: 11).

In the *interpersonal* domain, a DM can work to signal subjective functions. For instance, a speaker might use a DM to "express a response or reaction to the preceding or following discourse, to denote an attitude, to signal understanding or continued attendance, or to hedge an opinion" (Brinton, 2017: 11). In addition, a DM can signal intersubjective functions under the interpersonal domains. These functions include use of a DM to "effect cooperation or sharing, to show intimacy, to confirm shared assumptions or knowledge, to claim the attention of the hearer, to check understanding, to request confirmation, to express deference, or to address the positive or negative face" (Brinton, 2017: 11).

- Aijmer's (2002) classification:

As discussed in section 2.3.4, Aijmer's functional framework consists of two domains, textual and interpersonal, following earlier studies. According to her framework,

DMs in the textual domain work at a local or global level to signal frame or qualifier functions. At the local level (frame), Aijmer lists six functions defined in the following quotation:

1. Marking transitions (e.g. topic shifts, introducing a new aspect of the topic, opening and closing conversation)
2. Introducing a new turn (initiators)
3. Introducing an explanation, justification, background.
4. Introducing or closing a digression (push-markers, return-pops)
5. Self-correction
6. Introducing direct speech. (Aijmer, 2002: 42).

At the global level (qualifier), a DM indicates agreement/disagreement, a response to a question, a comparison/contrast, or a listing.

Meanwhile, Aijmer (2002) identifies the interpersonal domain as covering the phatic functions of DMs. These functions are related to politeness, hedging, planning, and showing intimacy.

- Müller's (2005) classification

Müller's (2005) model of DM functions follows previous research on DMs (e.g. Schiffrin, 1987; Brinton, 1996). However, she elaborates her list of functions to include individual functions for each DM discussed in her study. For instance, the DM *like* carries five functions defined in the following quotation:

1. searching for the appropriate expression
2. marking an approximate number or quantity
3. introducing an example
4. introducing an explanation
5. marking lexical focus (Müller, 2005: 246).

The other three DMs in Müller's study, namely *well*, *you know* and *so* carry both textual and interpersonal (*interactional* in Müller's terminology) functions. I will expand on her list of individual functions for these certain DMs in sections 4.5.1, 4.6.1 and 5.3.1, respectively.

Each of these scholars identifies general functions that multiple DMs may have, and then specifies certain subfunctions for the particular DMs that they investigate when they introduce their analyses of each particular DM. In my proposed framework, I opted to collect all the possible functions for the most studied DMs in the literature. Therefore, while some of the functions in my framework are applicable to most DMs, others are distinctive of one or only a few DMs. I list these subfunctions in my framework in table 3.5; examples of each DM function, drawn from the literature, are given in table 3.6.

3.5.2.2. The construction of the DM functional framework

In my proposed framework (table 3.5), I define two main functional domains, textual and interpersonal, following Aijmer (2002) and Brinton (1996, 2017). Within the textual domain, I propose five main functions: *initiator, elaborator, filler, transitional,* and *closure* (see table 3.5). The initiator function (proposed in Aijmer's 2002 framework) includes functions related to introducing (i) a new turn or (ii) a new topic. According to the literature, these two subfunctions are fulfilled by many DMs (including *so, well*, *yeah*, *okay*, and *you know*). I expand on the two subfunctions of the initiator function in sections 4.6.3, 5.3.3.1, 6.2.3.1 and 6.3.3.1. Additionally to these, I propose introduce new information as a third subfunction of initiator, because like the existing two subfunctions it relates to the introduction of *newness*. Typically, the function of flagging new information is mainly performed by DM *like*. In the literature, this function is covered under the label of *lexical focus* introducing new or given information (see further section 5.2.1 on different functions of

like). But I treat these two types of lexical focus DM *like* as two separate functions: introducing new information and focusing on given information, with the latter subfunction being classed under the elaborator function.

The second main textual function in my framework is the *elaborator* function. DMs with elaborator function often signal an upcoming extension to preceding utterances within the same turn, or in the prior turn. This extension could take the form of an explanation, a reformulation, a justification or an exemplification. These subfunctions are the most pervasive functions of DMs, and the literature confirms that many DMs may perform these subfunctions (see discussion in sections 4.5.1, 4.6.1, 4.7.1, 5.2.1, 5.3.1, 6.2.1, 6.3.1 and 6.4.1). Similarly, in my own data, all eight DMs under scrutiny prove able to perform the elaborator function. As mentioned earlier, I classify DM *like* used as a marker of given information as performing a subfunction of the elaborator function, because when they use *like* in this way, speakers in this function are trying to expand on previously expressed information by repeating it (see further section 5.2.3.2).

The third main function in the textual domain is the filler function. *Filler* refers to the function of a DM used while a speaker attempts to process an upcoming utterance or utterances. In fact, the label *filler* includes all functions that relate to the speaker's need to inform others about the linguistic structuring of their utterance. This structuring includes hesitation, planning, speaker's need to repair themselves or speaker's difficulty in finding the appropriate word (Dinkar et al., 2020; see also Levelt, 1995; Clark & Fox Tree, 2002). A DM produced with this function plays a significant role in signalling a speaker's attempt to plan their upcoming utterance and maintain the coherence of their discourse. I list three subfunctions under the filler function: searching for appropriate word, hesitation and planning. DMs with these subfunctions often co-occur with filled or unfilled pauses, providing further evidence of their nature as different types of delays for processing time. I

added self-repair as a subfunction of filler functions, although it is often unaccompanied by any type of pause, because a DM with self-repair function often shows that a speaker is still processing the upcoming utterances (in this case, processing their repair of their own utterance) (Levelt, 1995). The different filler functions are not confined to certain DMs. In the literature, different DMs are described as serving various filler functions under different labels (see Müller, 2005; Aijmer, 2002, 2013; Buysse 2009, 2012, 2017; Crible, 2017, 2018; among others). In my data, the filler subfunctions are fulfilled by all eight of the DMs I investigate (see sections 4.5.3, 4.6.3, 4.7.3, 5.2.3.3, 5.3.3.3, 6.2.3.2, 6.3.3.3 and 6.4.3.2).

The last function under the textual domain is the transitional function. A DM with one of the subfunctions of the transitional function signals a transition between parts of the discourse. This transition could take the form of a topic shift, topic resumption, digression or closure. A DM with one of these subfunctions therefore works as a link between two adjacent parts of the discourse. Among the DMs that can have these transitional functions are *so*, *yeah*, *okay*, and *well* (see among others Beach, 1993, 2020; Buysse, 2009, 2012, 2014, 2015; Beeching, 2016). I explore these subfunctions particularly in sections 4.5.3, 5.3.3.4, 6.2.3.3, 6.3.3.4 and 6.4.3.3. According to the literature, two additional subfunctions of the transitional label are linked with one specific DM, *so*, namely indicating a result or a summary (see among others Van Dijk, 1979; Schiffrin, 1987; Fraser, 1990). I expand on the discussion of these two subfunctions in section 5.3.1.

Within the interpersonal domain, I did not group the possible interpersonal functions under specific single-word labels akin to the textual domain. In my framework, I list eleven interpersonal functions (without defining subfunctions) that are discussed thoroughly in the literature on DMs (see section 2.3.4). All these functions signal a speaker's attempt to show cooperation and intimacy with other speakers' turns and to align with other speakers' requests or questions. Most of these functions (e.g. marking a speech act of request or question, marking agreement/disagreement…etc.) are described in the literature as being associated with many DMs, such as *well*, *so*, *okay*, and *you know* (see for example Schiffrin 1987, Fraser, Blakemore, 2002, 2006, Svartvik, 1980). However, some interpersonal functions are limited to certain DMs. For instance seeking confirmation is a function of DMs *you know* and *okay* (see section 4.6.4 and 6.2.4.2). Another function exclusive to *you know* is marking shared knowledge (see further section 4.6.4). The function of backchannel similarly is mainly linked to particular DMs, namely *okay*, *yeah* and *oh* (see section 6.2.4.4, 6.3.4.5, and 6.4.4.6).

Some of the interpersonal functions may appear at initial examination of their definitions to be textual functions. However, in all these cases, I have opted to classify such functions as part of the interpersonal domain for specific concrete reasons. For instance, Brinton (2006) lists holding/giving the floor as a textual function. However, other studies (e.g. Müller, 2005; Aijmer, 2002, 2013) place it in the interpersonal domain, because it works to signal a relationship between the speaker and the hearer. This latter group of authorities was my justification for going against Brinton and including this function under the interpersonal domain. A similar example is the conclusion function, found only with DM *so*. Buysse (2012: 1768) lists this function as an interpersonal function because it signals shared knowledge between speakers, a judgement which I opted to follow. However, I do acknowledge that stereotypically, introducing a conclusion is considered a textual function,

as it acts to structure the discourse; the particularly reasons for this decision are outlined at length in section 5.3.1. I further discuss the interpersonal functions in sections 4.5.4, 4.6.4, 4.7.4, 5.3.4, 6.2.4, 6.3.4 and 6.4.4.

The framework is designed to capture most of the functions of DMs discussed in the relevant literature (see section 2.4.3). However, debate exists in the literature as to whether some functions should be classified as DM functions or not; as such, those functions are not included in the framework. For instance, the function of *like* as an approximative marker (i.e. to flag approximation with numbers) is discussed as a DM function in a number of studies (e.g. Schouroup, 1985; Underhill, 1988; Andersen, 1998, 2001; Jucker & Smith, 1998; Fuller, 2003c; Schweinberger, 2014; D'Arcy, 2006; 2017). However, many other researchers argue for this function as a non-DM use of *like* because its absence affects the truth-conditionality of its host utterance (see section 5.2.1. for a detailed discussion on the functions of *like*). I followed the latter group of researchers and classified approximative *like* with numbers as a non-DM use of *like* (see table 3.4). This classification was done at an early stage before assigning the DM functions for *like*, i.e. during the DM and non-DM classification process.

Table 3.5. *Framework of categories of DM functions*

Domain	**Functions**	
Textual	**1. Initiators**	Introduce a new turn
		Introduce new topic
		Introduce new information
	2. Elaborators	Provide explanation, a reformulation, a justification or exemplification
		Focus on given information

	3.	**Fillers**	Assist a speaker in articulating their thought or search for appropriate expression Mark a repair of a speaker's speech Mark hesitation
	4.	**Transitional**	Mark topic shift Mark topic-resuming, or digression Indicate a result, a summary, a closure
Interpersonal			Mark a speech act such as request or question Express opinion Indicate a response or reaction on preceding or following discourse Check or express understanding Mark conclusion Mark shared knowledge Express agreement or disagreement Face-saving Turn management: hold or give the floor Seek or request confirmation Backchannel

In table 3.6, I present examples taken from my data for each function in the proposed framework. In the following chapters 4, 5 and 6, I explain in details the use of the eight DMs under scrutiny to fulfil these functions with more examples from the two corpora.

Table. 3.6. *Examples of DMs with each of the functions defined in the framework*

Domain	Function	Example
Textual	Introduce a new turn/ topic	(23) 084601: ***So,*** how was your day. 084602: fine/ 084601: did you eat your breakfast/ (0846/ Ar-Children)
	Introduce new information	(24) 084801: yeah. like the big thing, and then there is ***like*** a slide you can go through. (0848/ Ar-Children)
	Provide explanation	(25) S0416: Naruto it 's a Japanese anime an anime is like a carton in jap- in Japanese style ***so*** they do n't do as much of like real-life acting they do it in cartoons but it 's not like a kids ' cartoon (S3CK, BNC2014 children)
	Provide reformulation	(26) 24301: I have a brother. and he is, mm six years old. and I have a sister, like two months, (.) and I have , a bigger sister right here, uh and her age is fifteen, and me, I'm nine. ***I mean***. I 'm eleven and a half. (0234/ Ar-Children)
	Provide exemplification	(27) 102301: yeah, and also/ my vacation I spent watching movies, and k-dramas ***like***, the Korean uh movies and I also watched

		Vincenzo / Vincenzo Cassano , and it was literally , the best k-drama ever / to watch (1023/ Ar-Children)
	Focus on given information	(28) 110202: me I drew boys , in the iPad , ***like*** my iPad / (1102/ Ar-Children)
	Assist a speaker in articulating their thought or search for appropriate expression	(29) 082402: when I take from her something, she said, uh I know you take it. Give me it. and , uh she 's uh , like , uh get in my bed and , uh , ***like*** , uh the bear (0824/ Ar-Children)
	Mark a repair of a speaker's speech	(30) S0416: I do n't really remember anything else though ***oh*** I do I do remember a bit yeah that was okay that was okay (S3CK/ BNC2014 children)
	Mark hesitation	(31) 085302: mm (.) ***okay,*** (.) mm what do you wanna be/ when you grow, (0835/ Ar-Children)
	Mark topic shift	(32) 085301: mm oh what 's your least favourite food , the food that you hate the most, 085302: mm (.) broccoli. 085301: oh my god / me too. I hate vegetables. they're disgusting. ***well***, do you want to go, to the movies/ 085302: yeah, (0846/ Ar-Children)

	Mark topic-resuming, or digression	(33) 091701: yeah , she was n't able because she was, 091702: pregnant/ 091701: yeah. ***so*** the prince uh- the prince men/ the king 's men I mean , took the flower (0917/ Ar-Children)
	Indicate a result	(34) S0416: [..] and like you have the TV you have the TV for most of the evening ***so*** then we could n't use the TV and we 'd all just be up in my room with not much space (S8G6/ BNC2014 children)
	Indicate a closure	(35) 094502: okay, ***so,*** nice to meet you/ 094501: nice to meet you too. (0945/ Ar-Children)
Interpersonal	Mark a speech act such as request or question	(36) 084601***: so.*** how did you find him. 084602: my mum go out (0846/ Ar-Children)
	Express opinion	(37) S0417: there 's like huge aid going to poor countries that kind of thing through Christian organisations or S0416: >> ***well*** I think I think most religions have a good cause but I think (S8Q3/ BNC2014 children)

	Indicate a response or reaction on preceding or following discourse	(38) 094601: what 's your favourite colour, 094602: ***well*** I kind of like, mm green/ but it is n't like my favourite/ but I like it, (0946/ Ar-Children)
	Express understanding	(39) 090202: and skiing and ballet and gymnastics 090201: yeah , 090202: and, pretty much gym class. 090201: ***oh/*** gym class. you like all gym class. ***okay.*** (0902/Ar-Children)
	Mark conclusion	(40) 080202: okay/ mm what's your favourite colour/ 080201: I like black, white/ and/ maybe red and pink. 080202: ***so*** mainly shades, okay. 080201: mm, are you good at swimming/ (0902/Ar-Children)
	Mark shared knowledge	(41) 090202: I 'm more like of a blue / like pastel kind of sky blue-ish. that shade kind of blue. ***you know***. 090202: okay/ I love/ like, white blue, (0902/Ar-Children)
	Express agreement or disagreement	(42) S0525: they will not want to come anyway so it 's just

		S0622: >> I think you're right ***yeah*** mm they probably won't (SYJK/ BNC2014 children)
	Face-saving	(43) S0417: so you've got a lot to catch up on then got like an hour session I reckon you'll love being able to play the guitar when you're older it'll be a great skill S0416: ***well*** I can play the guitar (SYWB/ BNC2014 children)
	Turn management: hold or give the floor	(44) 83301: we want to talk about scary movie, want scary things happen to me. you want to share with me/ 083302: yes, 083301: ***okay***. we can start. 083302: no, you start. (0833/Ar-Children)
	Request confirmation	(45) 110202: I have step sisters ***, you know***/ 110201: you have step sisters, 110202: yeah , all my sisters , are steps (1102/Ar-Children)
	Backchannel	(46) 090202: […] he go slow/ and so like, when he- the guy tells him to go fast / he wo n't go fast. because he 's scared / 090201: ***yeah,*** 090202: he gots scared , so,

		090201: ***yeah/*** 090202: so like be. like, you have to trust the horse. 090201: ***yeah /*** 090202: you have to trust him. (0902/Ar-Children)

3.5.2.3.Checking the reliability of the DM functional framework

During the coding process, I used a separate document to gather examples of each function for use in detailed qualitative analysis (see figure 3.3). In many cases, I could not determine the function of a DM based on the few words of context visible on the downloaded concordance line. In these cases, I utilised the extended context function in CQPweb, exemplified in figure 3.4, to read more widely around the example.

S0655: since loads of people in my class get it spelled wrong

S0653: mm

S0655: chocolate cake

S0653: very sweet indeed

S0655: I still do n't know how to spell spaghetti carbonara **well** I think I might

S0653: spaghetti is erm complicated because it 's got a silent H in it

S0655: spaghetti

S0653: it 's spag

S0654: >>it 's er

Figure 3.4. *Extended context view for an example of* well *in the BNC2014 children subcorpus.*

As discussed in section 2.3.4, DMs are widely known to be multifunctional linguistic elements. According to Aijmer (2002), DMs' multifunctionality complicates drawing distinctions between their functions. Aijmer proposes looking at DMs as having primary and secondary functions. Whereas the primary function is often textual, the secondary should be interpersonal. In my study, I tried to assign one main function to each single occurrence of a DM. To achieve this, I had to establish criteria that I could use for distinguishing certain functions across the textual and interpersonal domains. I often relied upon the position of a DM within its turn (turn-initial, turn-medial and turn-final position), following earlier studies on DMs (e.g. Heritage, 1984, 2018; Aijmer, 2013; Lam, 2007; Buysse, 2012; Beeching, 2016). In the analysis of the Ar-Children corpus, I was able to look at factors of prosody to distinguish among some DM functions, following the work of Lam (2007), Huang (2011) and Diskin (2017). However, prosodic information was unavailable for the BNC2014 children subcorpus. Therefore, I had to determine additional criteria not reliant on prosody when analysing the latter dataset (see section 4.6.3 and section 5.3.3.1). Despite my efforts to establish clear criteria, assigning single functions to each example of a DM was not straightforward, since the textual and interpersonal domains interweave in many examples. I analysed the data looking for contradictory examples that may carry two possible functions. Then, I opt to assign a single main function for each of these examples following Aijmer's (2002) distinction of primary and secondary functions. This was often done with assistance from a second coder.

To evaluate the reliability of this framework of functions, the aforementioned second coder applied the category labels to a sample of 500 (out of 1,251, 40.0%) occurrences of DM *like* and all 622 occurrences of DM *so* in the BNC2014 children subcorpus. I provided this second coder with a document outlining the framework of functions, which functions were associated with each DM, and examples from the BNC2014 children subcorpus. In

follow-up meetings, the second coder and I discussed the ambiguity between certain functions. Overall, the two coders agreed in 86.9% of cases, that is 1,122 out of 1,873. I calculated the *Cohen's kappa* statistic for this exercise (Landis & Koch, 1977: 160). The Cohen's kappa value was 0.80, which indicates substantial agreement between coders (Landis & Koch, 1977: 165). The types of disagreements tended to be minor. Mostly, disagreement occurred with regard to cases that are not clear at first reading. Example (47) is of this nature. Here, my own coding classified the instance of *so* as a sequential marker that marks a connection to the preceding utterance. However, the second coder classified this instance of *so* as an explanation marker, because it explains the preceding utterance. After reconsidering this example, I decide that the second coder was correct, and changed my own analysis. I looked again at all examples I classified as "sequential" to see if they ought to actually be "explanation" and made changes if necessary. Overall, the level of disagreement was considered acceptable.

(47) S0655: it's a Star Wars legend

S0653: >>mm

S0655: ***so*** it's a legend in a legend because Star Wars is real it just happened a long time ago in a galaxy far far away (S7WR/ BNC2014 children)

I also recoded samples of each DM two weeks after I completed the initial coding. In this round, I looked at my first coding and considered whether I still agreed with that coding or not. In cases where I found a mismatch, I carried out a full recoding of the label and applied the necessary amendments. I agreed to my first coding in 98.9 % of cases, that is 554 out of 560 (all the DMs in the Ar-Children corpus). Then, I calculated the *Cohen's kappa* statistic which was 0.9, and indicated almost perfect agreement between coders (Landis & Koch, 1977: 165). This step was taken to ensure the consistency in the coding procedure.

3.5.3. Statistical analysis

Inferential statistical techniques that account for differences in speakers (e.g. regression models or ANOVAs) are commonly used in the field of CL. However, in the present study, I employed Log-Likelihood tests, which treat all the data in the corpus sections being compared single aggregates. I acknowledge that this type of test can therefore ignore variation between individual speakers. However, I considered these distributional tests to be appropriate for two main reasons: the objectives of the study and the distributional nature of my data. The main objective of the study is to compare the frequency of DMs and their functions in the speech of L2 children learning English with those of L1 speakers. Accordingly, I designed my research questions (see section 1.3) and constructed my corpus to examine speakers as a collective group rather than as individuals. Unlike ANOVAs or regression models, which test differences between participants, Log-Likelihood tests look at differences in the dataset as a whole. In addition, the frequencies of DMs and their functions are not evenly distributed across the two datasets. While some DMs are highly frequent, others are comparatively rare; this is also true of the various DM functions. The Log-Likelihood test is particularly suitable for data that exhibits such discrepancies, since it compares frequency distribution without any assumption of normality (Dunning, 1993:6).

I carried out a series of Log-Likelihood tests to determine whether differences in the frequencies of the eight terms as DMs are statistically significant. These tests were carried out using R. The log likelihood test accounts for variation in the relative frequency of an item across multiple corpora or subcorpora (Rayson, 2019: 322). The level of significance was set at 0.1% for this study because the analysis of DMs revealed a small number of occurrences for certain DM functions. Thus, a small threshold can ensure that meaningful patterns are not overlooked, particularly in cases where detecting subtle effects is critical to understanding the functions of these DMs.

Additionally, since some DM frequencies are low (see section 4.5), performing tests on each individual marker would not yield meaningful results. Therefore, I instead opted to look at the frequencies of all eight DMs collectively, as a profile of DM usage for each corpus, and to test the significance of the difference between the two profiles. The results of these tests are presented and discussed in subsequent chapters, according to each DM. Significance tests of this kind need to be based on the number of times the opportunity to use a DM arises and is taken, compared with the number of times an alternative is used (i.e. an opportunity arises but is not taken). These are also the numbers needed to calculate relative frequency (see section 2.4.4.3).

Determining the number of opportunities for a DM to be used is not straightforward. I followed the procedure utilised by Stefanowitsch & Gries (2003: 213). These authors treat the overall number of clauses in each (sub)corpus under examination as the number of opportunities for the phenomenon under investigation to occur. This approach is predicated on the assumption that a clause is highly unlikely to contain (or be associated with) more than one instance of the phenomenon being studied. While Stefanowitsch & Gries do not look at DMs, I consider it a reasonably safe approximation to assume that in normal discourse, a single clause provides one opportunity to use a DM. Although it is common in English generally for multiple DMs to cluster within a single clause, this phenomenon occurs only rarely in my data. This indicates that DM clustering does not invalidate the usefulness of counts of clauses as the basis of normalisation for the frequencies of DMs and DM functions.

Stefanowitsch & Gries suggest that a clause can be defined by the presence of a main verb, so a count of main verbs in a corpus is a good proxy for a count of clauses. Thus, in order to estimate the number of clauses in my two corpora, I performed a series of simple corpus queries using POS tags (see table 3.7). First, I counted the number of lexical verbs in each (sub)corpus using query (a) from table 3.7. These lexical verb lemmas always function

as main verbs. Conversely, modal verbs could be ignored for these purposes, because they are always auxiliaries. However, this count ignores the possibility of the primary auxiliary verbs *be*, *do* and *have* occurring as main verbs. To add to the total the number of clauses in which these three verbs function as main verbs and not auxiliaries, I performed a series of calculations based on additional queries. First, I retrieved and counted all the occurrences of the primary auxiliaries in the corpus (using query (b) in table 3.7). From this overall frequency, I subtracted the count of cases in which the instance of *be*, *have* or *do* is indeed an auxiliary. These were identified using queries (c + d) in table 3.7, which find auxiliary-usage *be*, *have* and *do* in declarative and interrogative clauses respectively. The estimated numbers of clauses could then be calculated from the counts of results for each of queries (a), (b), (c) and (d), as follows:

Total estimated number of clauses = a + (b - c - d)

Table 3.7. *Simple queries for types of verb in CQPweb using CLAWS tags*

Label	Query target	Query
a	**Lexical verbs**	_VV*
b	**All primary auxiliaries**	_V[B,H,D]*
c	**Primary auxiliaries with auxiliary function in declarative clauses**	_V[B,H,D]*(_[R,X]*)*_V*
d	**Primary auxiliaries with auxiliary function in interrogative clauses**	_V[B,H,D]*(_X*)?(_P*|(_[A,D]*)?_N*)_V*

Via this process, I estimated that there are 10,729 clauses in the Ar-Children corpus and 22,627 clauses in the BNC2014 children subcorpus (see table 3.8). This allowed me to estimate the number of clauses that include a particular DM based on that DM's overall frequency (and likewise the number of clauses that do not), and to perform subsequent statistical comparisons.

Table 3.8. *Number of clauses in the two corpora*

Label	**Query target**	**Ar-Children corpus**	**BNC2014 children subcorpus**
a	**Lexical verbs**	7,191	13,684
b	**All primary auxiliaries**	5,170	13,317
c	**Primary auxiliaries with auxiliary function in declarative clauses**	1,049	3,512
d	**Primary auxiliaries with auxiliary function in interrogative clauses**	583	862
Estimated N of clauses	**a + (b – c – d)**	10,729	22,627

I employed the same approach to calculate estimated counts of clauses for the six subcorpora compiled from the Ar-Children corpus (see section 3.4), as presented in table 3.9.

Table 3.9. *Number of clauses in the six subcorpora of Ar-Children*

Label		Arabic-speaking preschool		English-speaking preschool			
	Age	9-10	11-12		9-10		11-12
	Years in English-speaking countries	0	0	1+	2+	1+	2+
a	**Lexical verbs**	1,623	1,134	1,074	1,022	724	426
b	**All primary auxiliaries**	1,087	890	509	438	360	241
c	**Primary auxiliaries with auxiliary function in declarative clauses**	219	178	72	53	61	53
d	**Primary auxiliaries with auxiliary function in interrogative clauses**	111	114	131	157	122	58
Estimated N of clauses	**a + (b – c – d)**	**2,380**	**1,732**	**1,583**	**1,460**	**1,084**	**667**

Calculating relative frequencies of DMs on the basis of clause counts has the advantage that it does not risk overstating statistical significance by calculating the log likelihood using the huge numbers from counts of word tokens, which is the usual default basis of normalisation. However, doing this has the possible disadvantage that the basis of normalisation may not perfectly reflect the underlying distribution. To address this possibility, I recalculated my comparisons using word token counts as the basis of

normalisation. The results of the differences among the relative frequencies in the recalculation were similar to those that I originally obtained.

3.6. Summary

This chapter has outlined the methodology of the thesis. In section 3.2, I explained in detail the process of compiling the Ar-Children corpus, covering the recruitment of participants, the ethical considerations and the process of recording and transcribing the conversations. The Ar-Children corpus consists of 60 texts, containing speech by 120 female participants recorded in semi-spontaneous conversations between peers. In section 3.3, I discussed the construction of a comparable dataset of L1 school-age children as a subcorpus of the Spoken BNC2014. I described the process of defining a number of subcorpora of the Ar-Children corpus in section 3.4. The purpose of constructing these subcorpora is to evaluate the effect of external factors (context and duration of preschool attendance) on the acquisition of DMs by the participants whose speech is contained in the Ar-Children corpus.

In section 3.5, I discussed the analysis procedure. The DMs to be analysed were identified by examining wordlists for each corpus. I then outlined the concordance coding process, and proposed a framework of DM functions based on key literature. I then explained the statistical procedures used to measure relative frequencies of the different items and functions in each of the (sub) corpora, and to test any differences in frequency for statistical significance in the two corpora with justification for choosing these types of test and method of calculation. The following chapters present a quantitative analysis of DM frequencies as well as a qualitative analysis of DM functions in the two corpora. The analyses across chapters 4, 5 and 6 highlight observable similarities and differences between the use of these markers in the speech of L1 and L2 speakers.

Chapter 4: Overview of DM frequencies and functions

4.1. Overview

This chapter presents an overview of the results regarding DMs in the Ar-Children corpus and the BNC2014 children subcorpus. It begins with a discussion of DM frequencies in the two corpora. These results demonstrate that the DMs *well*, *you know* and *I mean* occur with relatively low frequency in the Ar-Children corpus; for this reason, the qualitative aspects of these DMs, in terms of their functions in context, are explored within this chapter. The analysis of the other five, more frequent DMs warrant more extended investigation and so will be presented in subsequent chapters.

In section 4.2, I answer RQ1: *are there any differences between the frequencies of the most frequent DMs in the ESL speech of Arabic-speaking upper-grade children (fourth, fifth and sixth grades) in Saudi international primary schools, compared to age-matched native English-speaking children?* To answer this question, I compare the frequencies of the eight DMs across the two corpora. In section 4.3, I present the answer to RQ2: *how does the context of formal exposure to English prior to joining primary school affect the frequency of DMs in the ESL speech of these Arabic-speaking children?* This question is predicated in the hypothesis that the frequency of DMs will be higher in the speech of children who have attended preschool in English-speaking countries than in the speech of those who spent their preschool years in Arabic-speaking countries. In addition, it is predicted that the longer these children have been exposed to English in English-speaking preschool, the more frequent their use of DMs will be. To answer this question, I analyse DM frequencies in relation to two of the metadata variables in the Ar-Children corpus, namely, the context of preschool study (in an English-speaking country or an Arabic-speaking country: see section 3.4) and the length of time spent in English-speaking preschool.

In section 4.4, I then compare DM frequencies between the two corpora, according to the textual and interpersonal functional domains. This section answers the third research question: *what are the functions of the most frequent DMs in the speech of these Arabic-speaking school-age children? Are there any differences in the functions with which DMs are used, in comparison to native English-speaking children?* Answering this question involves a jointly quantitative and qualitative analysis of the data. The quantitative part is presented in section 4.4 in its entirety, but the qualitative analysis begins in section 4.5 and continues through subsequent sections and chapters. In the remainder of this chapter, sections 4.6 to 4.8, a qualitative analysis of the DMs *well, you know* and *I mean* will be presented; analysis of the other five DMs is presented in chapters 5 and 6.

4.2. Overall frequencies of DMs in the two corpora

This section will answer RQ1, comparing the frequencies of the eight DMs in the Ar-Children corpus and the BNC2014 children subcorpus. As established in section 3.4, the DMs under study are *like, yeah, okay, oh, well, you know* and *I mean.* Table 4.1 presents the overall raw and relative frequencies of these eight forms in the Ar-Children corpus and the BNC2014 children subcorpus. The relative numbers are normalised on a basis of 100 clauses (see section 3.5.3). These figures precede my division of the examples into actual DMs versus the same forms used with non-DM functions.

Table 4.1. *Raw & relative frequencies (per 100 clauses) of the forms of DMs in the Ar-Children corpus and the BNC2014 children subcorpus*

Expression	Ar-Children corpus		BNC2014 children subcorpus	
	N	Per 100 clauses	N	Per 100 clauses
Like	2,069	26.5	2,306	17.2
So	642	8.2	952	7.1
yeah	848	7.9	2,653	11.7
okay	560	5.2	448	2.0
oh	317	4.1	1,068	4.8
you know	181	2.3	195	1.5
well	58	0.8	805	6.0
I mean	25	0.3	135	1.0

Like is the most frequent of the eight forms in the Ar-Children corpus, but second most frequent in the BNC2014 children subcorpus. *Yeah* and *oh* occur more frequently in the BNC2014 children subcorpus than in the Ar-Children corpus. Conversely, *okay* and *you know* are twice as frequent in the Ar-Children corpus as in the BNC2014 children subcorpus. *I mean* is three times more frequent in the BNC2014 children subcorpus, and most notably, *well* is over seven times more frequent in the BNC2014 children subcorpus than in the Ar-Children corpus.

However, the frequencies in table 4.1 do not distinguish DM and non-DM functions of the forms in question. They thus only give a general indication of DM frequency. As explained in section 3.5.1, I separated DMs from non-DM instances by manually inspecting all the concordance lines for each DM in both corpora. Table 4.2 presents the resulting

frequencies of the eight forms as DMs and as non-DMs. The numbers in table 4.2 are more informative as they give the precise frequency of each form as a DM.

Table 4.2. *Raw frequencies and percentages for the eight forms as DMs and non-DMs in the Ar-Children corpus and BNC2014 children subcorpus.*

Expression	**Ar-Children corpus**		**BNC2014 children subcorpus**	
	DM use (%)	Non-DM use (%)	DM use (%)	Non-DM use (%)
like	1,199 (57.4)	870 (42.6)	1,251 (54.2)	1,055 (45.8)
So	322 (50.2)	320 (49.8)	622 (65.3)	330 (34.7)
yeah	639 (74.9)	209 (25.1)	1,668 (62.9)	985 (37.1)
okay	548 (97.9)	12 (2.1)	397 (88.7)	51 (11.3)
oh	294 (92.7)	23 (7.3)	1,029 (96.3)	39 (3.7)
you know	86 (53.6)	85 (46.4)	114 (58.5)	81 (41.5)
well	46 (79.3)	12 (20.7)	704 (87.5)	101 (12.5)
I mean	25 (100.0)	0 (0.00)	114 (84.4)	21 (15.6)
total	3,159 (65.3)	1,550 (34.7)	5,899 (68.9)	2,663 (31.1)

Like and *yeah* are the most frequent DMs in both corpora, just as they were the most frequent of the eight forms. The majority of instances of *oh*, *okay* and *well* are DMs in both corpora. The low frequency of *I mean* means that there are no non-DM examples in the Ar-Children corpus. Finally, in both corpora, more than half the occurrences of *so* and *you know* function as DMs. The next step was a comparison across the two corpora of the frequencies of these forms as DMs. Figure 4.1 shows relative frequencies for each DM in the two corpora, normalised to frequency per 100 clauses.

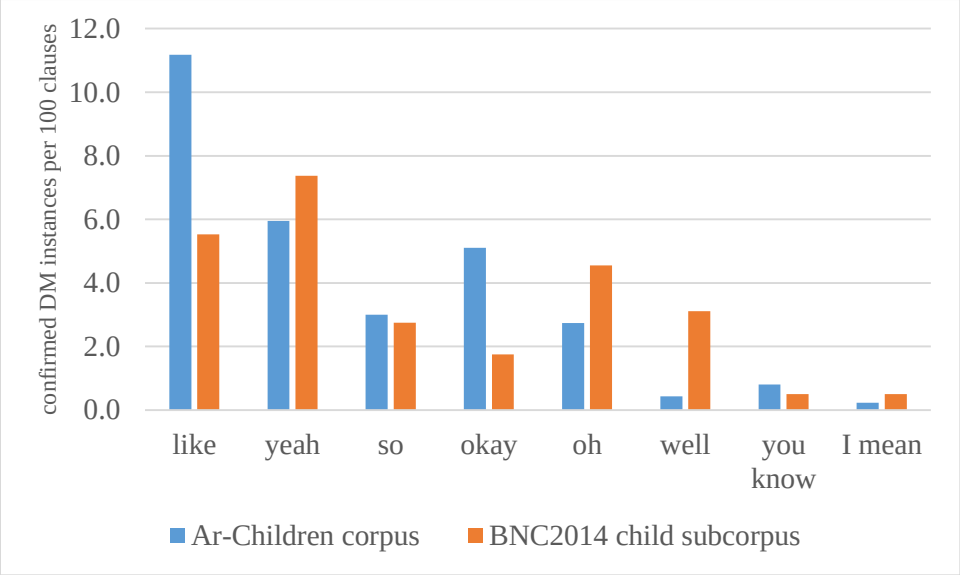


Figure 4.1. *Relative frequencies of confirmed DMs per 100 clauses, calculated from their raw frequencies as presented in table 4.2*

Like and *okay* are more frequent in the Ar-Children corpus (11.2 and 5.1 per 100 clauses, respectively) than the BNC2014 children corpus (5.5 and 1.8 per 100 clauses, respectively). Conversely, *yeah* and *oh* are more frequent in the BNC2014 children subcorpus (7.4 and 4.5 per 100 clauses, respectively) than in the Ar-Children corpus (6.0 and 2.7 per 100 clauses, respectively). While *I mean* is one of the least frequently used DMs in both corpora, it is about three times more frequent in the BNC2014 children subcorpus (0.5 per 100 clauses) than in the Ar-Children corpus (0.2 per 100 clauses). The frequency of *so* and *you know* as DMs appears to be fairly similar across both corpora. These results are consistent with the overall picture evident in table 4.1. This may be attributed to the fact that all eight forms function mostly as DMs in the two corpora (see table 4.2).

A Log-Likelihood test (LL) was conducted to determine whether the observed difference in the overall pattern of frequency of the eight DMs between the Ar-Children corpus and the BNC2014 children subcorpus is statistically significant (see section 3.5.3). To perform the test, I created a contingency table of two columns corresponding to the two

corpora. Each column has nine rows, containing the eight DM frequencies for that corpus and the frequency of clauses with no DMs. This Log-Likelihood test resulted in a *p* value = $2.2x10^{-16}$, LL = 971.4 and df = 8. The difference in frequency profiles between the two corpora is therefore statistically significant. Having established that the overall profiles are significantly different, the rest of the analysis will explore those differences in greater detail, probing issues such as whether particular DMs are being used for different or similar functions across the two corpora.

The most notable similarity between the corpora is the relatively low frequency in both of the two multiword DMs *you know* and *I mean*. Many studies report similarly low frequencies for *you know* and *I mean* in child speech (e.g. Andersen, 2001; Macaulay, 2002, Romero-Trillo, 2002; Beeching, 2016). A possible interpretation of these DMs' low frequency is that children in the age range represented in my data are still in the process of acquiring these forms. I discuss the relationship of all my findings to the literature on DM acquisition further in section 7.3.

Since the frequencies of *well*, *you know* and *I mean* are low, in the Ar-Children corpus in particular, they will be discussed only briefly, in sections 4.5, 4.6 and 4.7.

4.3. Analysis of DM frequency in the Ar-Children corpus according to time spent in preschool

In this section, I consider the frequency of DMs in the Ar-Children corpus in relation to number of years each speaker has spent in English-speaking countries during pre-school (i.e., before primary school). This analysis answers the second research question: how does the context of formal exposure to English prior to joining primary school affect the frequency of DMs in the ESL speech of these Arabic-speaking children? The question is divided into two sub questions:

a) Is there any difference in the frequencies of DMs between participants who have studied ESL in preschool in Arabic-speaking versus English-speaking countries?
b) Is there any effect of the number of years of attendance at preschool in English-speaking countries on the frequencies of DMs?

The questions are predicated on two hypotheses. First, it is hypothesised that the frequency of DMs will be higher in the speech of children who have spent their preschool years in English-speaking countries, compared to preschools in Arab countries. Second, it is hypothesised that, among the former group, the longer these children have been exposed to English in English-speaking countries, the more frequent their use of DMs will be. These hypotheses are based on previous literature that report positive correlation between time spent in English-speaking countries and the acquisition of DMs (see 2.3.5). The duration and context of formal exposure to the L2 (e.g. in L1 context vs. abroad/L2 context) are among the factors that may determine the amount of L2 input a child receives and thus may affect their language development (see section 2.2.1). As such, it is important to account for these factors when studying child SLA (Paradis, 2011: 213).

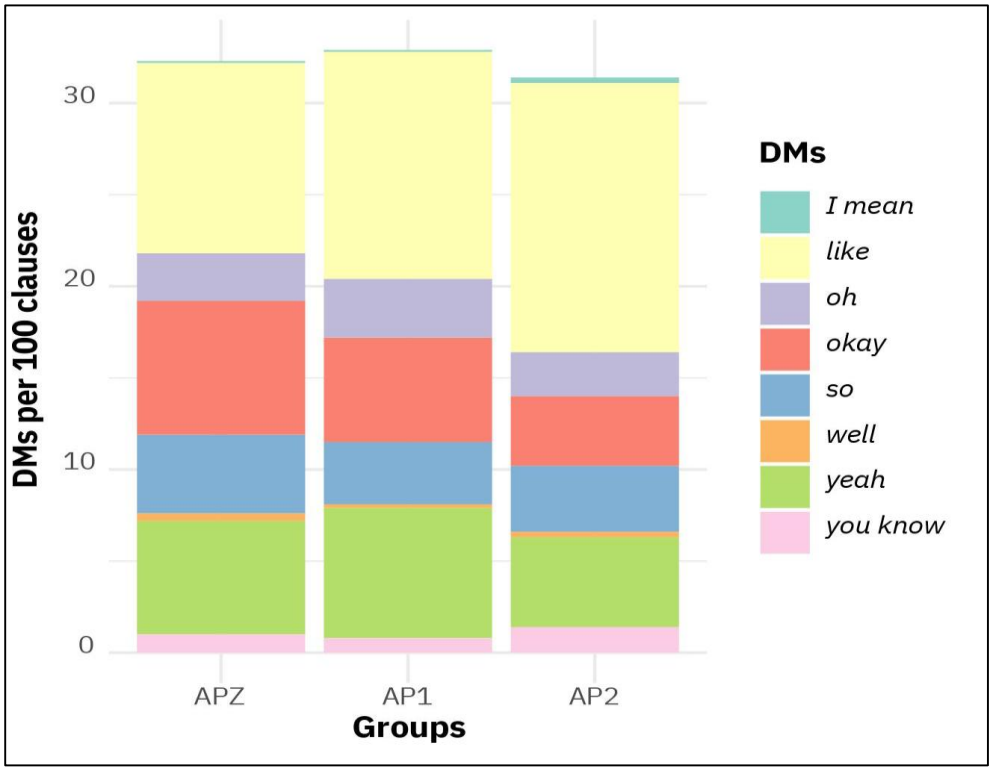
To derive the necessary frequency data, I counted the instances of DMs per individual speaker (see Appendix F for the resulting table). At this point, frequencies were calculated for the groups of speakers whose definition was discussed in section 3.4.

Table 4.3 presents raw and relative (per 100 clauses) frequencies of DMs for the 9 to 10 year olds; this data is visualised in figure 4.2.

Table 4.3. *Raw and relative (per 100 clause) frequencies of DMs in the speech of 9-10 year olds*

Years		*like*	*so*	*well*	*yeah*	*okay*	*oh*	*you know*	*I mean*	Total
APZ	N	248	102	9	148	173	62	23	3	**768**
	RF	10.4	4.3	0.4	6.2	7.3	2.6	1.0	0.1	**32.3**
	%	32.3	13.3	1.2	19.3	22.5	8.1	2.9	0.4	**100**
AP1	N	196	54	3	112	90	50	13	2	**520**
	RF	12.4	3.4	0.2	7.1	5.7	3.2	0.8	0.1	**32.9**
	%	37.7	10.4	0.6	21.5	17.3	9.6	2.5	0.4	**100**
AP2	N	214	52	4	72	56	35	21	5	**459**
	RF	14.7	3.6	0.3	4.9	3.8	2.4	1.4	0.3	**31.4**
	%	46.6	11.3	0.9	15.7	12.2	7.6	4.6	1.1	**100**

Figure 4.2. *Relative frequencies (per 100 clauses) of DMs in the speech of 9-10 year olds*



The overall frequencies of DMs are comparable across the three groups (32.3 per 100 clauses for the APZ group, 32.9 for AP1 and 31.4 for AP2). However, the frequency of DMs becomes lower with more time spent in preschool in English-speaking countries (AP1 vs. AP2). Looking at individual DMs, *yeah*, *oh* and *okay* become less frequent with more time spent in English-speaking countries, but on the other hand, *like*, *so*, *you know* and *I mean* are used more frequently by the AP2 group.

This comparison shows that experience of preschool in English-speaking countries has no clear positive effect on the frequency of DMs. In fact, speakers who attended preschool in Arab countries use DMs with similar frequency to those who attended preschool abroad.

Table 4.4 present the equivalent data for the 11 to 12 year olds; the data is visualized in figure 4.3.

Table 4.4. *Raw and relative frequencies per 100 clauses and percentages of DMs in the speech of 11-12 year olds*

Years		***like***	***so***	***well***	***yeah***	***okay***	***oh***	***you know***	***I mean***	**Total**
APZ	N	210	40	1	131	107	37	4	5	**535**
	RF	12.1	2.3	0.1	7.6	6.2	2.1	0.2	0.3	**30.9**
	%	39.3	7.5	0.2	24.5	20.0	6.9	0.7	0.9	**100**
AP1	N	193	24	9	78	22	34	18	2	**380**
	RF	17.8	2.2	0.8	7.2	2.0	3.1	1.7	0.2	**35.1**
	%	50.8	6.3	2.4	20.5	5.8	8.9	4.7	0.5	**100**
AP2	N	62	20	11	67	31	48	1	6	**246**
	RF	9.3	3.0	1.6	10.0	4.6	7.2	0.1	0.9	**36.7**
	%	25.2	8.1	4.5	27.2	12.6	19.5	0.4	2.5	**100**

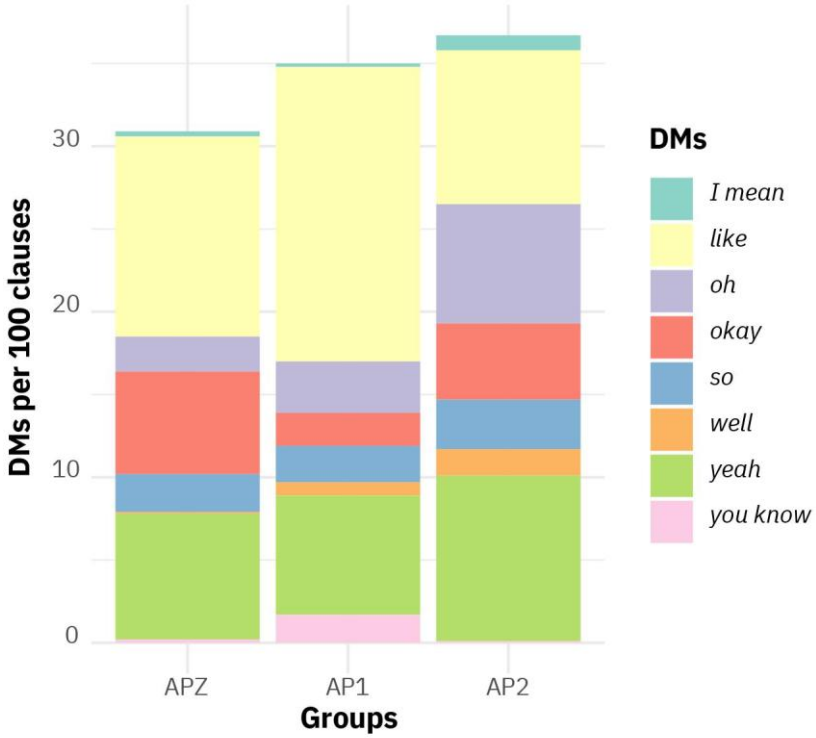


Figure 4.3. *Relative frequencies (per 100 clauses) of DMs in the speech of 11-12 year old*

For the children aged 11-12, DM frequencies are higher among those who have spent more time in preschool in English-speaking countries, as Figure 4.3 clearly shows. The overall frequency of DMs is highest in the AP2 group. If we consider the different DMs separately, we can see that AP2 relies less on *like* –using *yeah, well, oh* and *so* more frequently. *You know* is more frequent in the AP1 group than the other groups. The APZ group employs *okay* more frequently than the other groups. I will discuss this further in section 7.4.

A Log-Likelihood (LL) test was conducted to test whether the observed differences in the pattern of frequency of the eight DMs between the subcorpus of students who attended English-speaking-country preschool and the subcorpus of students who attended the Arab-country preschools is statistically significant. To perform the test, I created a contingency

table that consists of two columns, one for APZ and one for AP1 and AP2 combined. Each column has nine rows consisting of the eight DM frequencies in the subcorpus and the frequency of clauses with no DMs. The test was performed for each age group separately. The difference in frequency profiles between the APZ and AP1/AP2 subcorpora of 9-10 year olds is statistically significant, with a *p*-value less than 2.2×10^{-16}, LL= 310.4 and df = 16. The same test applied to the data for 11-12 year olds resulted in a *p*-value less than 2.2×10^{-16}, LL= 301.23 and df = 16. This shows that there is a difference between the three groups in their DM use.

The results in this section show that the frequencies of the eight DMs do not increase in step with the children's number of years of preschool in English-speaking countries. The 9-10 year-old group exhibit *fewer* DMs the more English-speaking-country preschool experience they had. This result contradicts prior studies that found evidence of the positive effect on the frequency of DMs from exposure to English as a L2 in English-speaking contexts in adult L2 speech (Hellermann & Vergun, 2007; Caliendo & Magliacane, 2020; Diskin-Holdaway, 2021; Huang et al., 2023 among others). Rather, this result suggests that spending time in the L1 context prior to primary school does not result in more frequent use of DMs five or seven years later. However, although speakers in the APZ category used DMs almost as frequently as the other groups, they tended to use a less varied spread of DM types. Speakers in the APZ group, in both age groups, used *okay* more than the other DMs to a degree not observed for AP1 and AP2 (22.5 % in 9-10 year olds and 20% in the 11-12 year old group). On the other hand, *you know* and *I mean* are used more frequently among participants in the AP2 group. The *variety* of DMs is higher with more time spent in an English-speaking-country preschool in both age groups. Furthermore, AP2 speakers use more varied DMs than AP1 in both age groups. This shows that Arabic-speaking children who attended preschool in English-speaking countries have more varied usage of DMs than have

those who did not spend any time in an English-speaking preschool. This result is in line with prior studies on adult learners that reported a higher variety of DMs in the speech of more experienced English learners compared to less experienced ones (e.g. Liu, 2013, 2016; Ament & Pérez, 2018; Kwon, 2022). Combining the results presented here and previous studies, we could conclude that learners with more experience of English in English-speaking countries are able to use DMs in a more varied manner. I will discuss this further in chapter 7.

4.4. Quantitative analysis of DM functions across the two corpora

As explained in section 3.6, I manually coded each DM for function, with particular attention to the functional domains called textual and interpersonal. The textual domain includes the functions of DMs that assist speakers in managing discourse and achieving coherence. The interpersonal domain includes the functions that assist speakers in turn management and in expressing emotions and attitudes. Table 4.5 shows the raw frequencies (and percentages) of each DM in the Ar-Children corpus and the BNC2014 children subcorpus being utilised within each of the two functional domains. Unclassified or unclear DMs are excluded here (see section 3.5.1). The specific textual and interpersonal (sub)functions in question will be detailed in the subsequent sections dedicated to each.

Table 4.5. *Raw frequencies and percentages for domains of DM function in the Ar-Children corpus and BNC2014 children subcorpus*.

DMs	Ar-Children corpus				BNC2014 children subcorpus			
	Textual		Interpersonal		Textual		Interpersonal	
	N	%	N	%	N	%	N	%
like	1,176	100	0	0.0	1,251	100	0	0.0
So	257	79.8	65	20.2	387	63.3	224	36.7
yeah	72	11.3	567	88.7	128	8.2	1,440	91.8
okay	237	43.2	311	56.8	41	10.8	339	89.2
Oh	83	28.2	211	71.8	338	36.5	589	63.5
well	22	45.8	26	54.2	365	52.1	327	47.9
you know	51	59.3	35	40.7	62	61.4	39	38.6
I mean	19	76.0	6	24.0	77	77.8	22	22.2
total	1,917	61.1	1,221	38.9	2,649	47.1	2,980	52.9

Like is almost always a textual marker in both corpora[6] (see further section 5.2). Similarly, *so*, *you know* and *I mean* carry mainly textual functions in both corpora (see further sections 5.3, 4.6 and 4.7). By contrast, *yeah* and *okay*, both of which have semantics of agreement, function mainly as interpersonal DMs in both corpora (see further sections 6.4 and 6.2). More than two thirds of instances of *oh* function within the interpersonal domain in both corpora (see further section 6.3). *Well* is mainly an interpersonal DM in the Ar-Children corpus, but mostly a textual DM in the BNC2014 children subcorpus (see further section 4.5).

[6] Some interpersonal functions were observed for *like* but not coded formally. I discuss thoroughly the reasons for not coding these instances as interpersonal functions in section 5.2.4.

A series of Log-Likelihood tests were conducted to test whether the observed differences in textual versus interpersonal functions are statistically significant. A test was performed for each DM separately. I created a two-by-two contingency table with columns corresponding to the two corpora. Each column has two rows containing the frequencies of the DM being used with textual and with interpersonal function. The results of the Log-Likelihood tests are presented in table 4.6. The differences between the two corpora are statistically significant for *so* and *okay* at the $p < 0.001$ level. For the other DMs, there were no significant differences at the $p < 0.001$ level. (In all these tests, df =1)

Table 4.6. *Log-Likelihood test results for differences in frequency of DM functions between the two corpora*

DMs	**LL**	***P*-value**
So	27.95	1.25×10^{-7}
Okay	123.43	2.2×10^{-16}
Oh	6.85	0.01
Yeah	5.11	0.02
Well	0.86	0.3541
You know	0.08	0.7715
I mean	0.04	0.8501
Like	0	1

We see, then, that there is no significance difference in the use of six out of the eight DMs in the speech of Arabic-speaking children compared to L1 children in terms of preference for textual versus interpersonal functions. This finding is in line with Lam's (2007) study. Lam reports finding no difference in the functions of *well* and *so* between L1 and L2 adult speakers. However, many other studies have found that L1 speakers tend to use DMs to serve interpersonal functions, whereas L2 speakers employ DMs mainly with textual

functions (e.g. Müller, 2005; Buysse, 2012; Jakupčević, 2019; Özturk & Köse, 2021; Pan & Aroonmanakun, 2022). Counter to these studies, my finding of a high frequency for the interpersonal function in the Ar-Children corpus reveals that these L2 speakers, at least, are in fact able to use DMs with interpersonal functions. I discuss this further in section 7.3.

This section has presented the frequencies of DM functions according to the two broad functional domains of textual function and interpersonal function. These frequencies guide the manner in which I opt to present the subsequent phases of this study. The DMs *like* and *so* have predominantly textual functions; subsequently, they are the subject of Chapter 5. The DMs *yeah*, *okay* and *oh* are principally interpersonal DMs; as such, they are the subject of Chapter 6. The remaining three DMs under study occur with much lower frequency and thus merit less extensive discussion. For that reason, *well*, *you know* and *I mean* are addressed briefly in the remaining sections of the current chapter.

4.5. *Well*

4.5.1. *Well* in previous literature

Well is one of the most studied DMs (e.g. Lakoff, 1973, Svartvik, 1980; Owen, 1981; Carlson, 1984; Schouroup, 1985, 2001; Schiffrin, 1987; Watts, 1989; Norrick, 2001; Müller, 2004; Lam, 2007; Aijmer, 2002, 2011, 2013, 2015; Heritage, 2015; Huang, 2019). Svartvik (1980) offers the first comprehensive analysis of *well*, analysing examples from the London-Lund Corpus of Spoken English. According to Svartvik (1980: 173), *well* functions as either a *qualifier* or a *frame*, terms corresponding to what I refer to as the textual and interpersonal domains respectively. As a textual DM (or frame in Svartvik's terms), *well* links adjacent discourse units by initiating or closing a section of discourse, or marking an explanation or a clarification (Svartvik, 1980: 173; see also Carter & McCarthy, 2006: 214, Beeching, 2016: 185; Aijmer, 2013: 34). It can also mark self-repair, i.e., it can preface a correction or

modification of a previously uttered answer or comment (Svartvik, 1980: 175; see also Schouroup, 1985: 56; Aijmer, 2013: 32).

Well as a textual DM can be an editing marker – that is, an element that signals planning (Aijmer, 2011: 237). This editing function is also characterised as the speaker searching for the right phrase (Müller, 2005), or as a processing function (Lam, 2007). While searching for a phrase or a word, speakers use *well* as a "delaying tactic" to gain time to plan their speech (Svartvik, 1980: 171; see also Jucker, 1993: 448; Huang, 2019: 581). Aijmer (2013: 44) argues that *well* appears in turn-medial position, with filled or unfilled pauses, to mark coherence editing functions (filler functions in my framework, see section 3.5.2). In fact, Svartvik (1980) points out that *well* as a frame (textual in my framework) often appears in turn-medial position.

Well can also flag a shift to reported speech (Svartvik, 1980: 173; Schouroup, 1985: 18, Schiffrin, 1987: 102). This function occurs with quotative verbs such as *say, think, feel* or *be+like*; hence, it is labelled by Müller (2005:113) as *quotative well*. Svartvik (1980: 175) claims that this function of *well* as a direct speech marker in spoken language is similar to quotation marks in writing. Although quotative *well* is discussed thoroughly in the literature (e.g. Svartvik, 1980; Schiffrin, 1987; Jucker,1997; Huang, 2019; Aijmer, 2013; Beeching, 2016), researchers have found that it is one of the least frequent functions of *well* as a DM (see Müller, 2005: 115; Buysse, 2015: 77).

The interpersonal functions of *well* (or qualifier in Svartvik's terms) include its use as a marker used by a speaker to take the conversational floor and introduce a new turn; when so used, it occurs in turn-initial position. In this function, speakers use *well* to alert other speakers about their intention to say something or comment on their speech (Svartvik, 1980: 175). In this function, speakers also use *well* to express emotional affect or soften their

comments (Huang, 2019: 582). Another interpersonal function of *well* is marking answers as insufficient (Svartvik, 1980: 172). This function was noted in an earlier analysis of *well* by Lakoff (1973), who observes that a speaker uses *well* to signal insufficiency in a direct answer to a question when they do not provide the required information explicitly. Such an insufficient answer often leads to another question (Lakoff, 1973: 460). In addition, *well* can precede an indirect answer to express insufficiency if the speaker believes the questioner will need to deduce the requested information from the answer given ("what time is it? *Well*, I just told Bill it was noon", Lakoff, 1973: 457).

The use of *well* to indicate insufficiency is also discussed by Schiffrin (1987:102). Schiffrin claims that *well* marks indirect responses to questions when the upcoming discourse "is not fully consonant with prior coherence options". Being "not fully consonant" means that the answer provided by the speaker is a dispreferred answer (Biber et al., 1998: 1068) or not a straightforward one (Schegloff & Lerner, 2009:101). These views are akin to Lakoff's (1973). On the other hand, Owen (1981:110) criticizes Lakoff's account of *well* as marking insufficiency in responses to questions. According to Owen (1981: 112), this use of *well* in fact functions to reduce the threat of losing face when answering questions or requests.

Jucker (1993: 444) agrees that *well* is employed with answers where there is some problem at the interpersonal level as a "face-threat mitigator". Huang (2011: 158) adds that *well* serves as a softener for the force of speech when it precedes a disagreement. In Aijmer's (2013: 40) characterisation, *well* as a disagreement marker is used when something does not meet the speaker's expectation (Aijmer, 2013: 40). But *well* does not usually occur with direct disagreement utterances unless the speaker wants to express politeness or dramatic effect (Carlson, 1984: 41).

In the field of SLA, researchers have observed instances in which L2 speakers use *well* differently from L1 speakers. For instance, Müller's (2005) results show that, relative to L1 speakers, German learners employ *well* more frequently as a marker of planning, i.e. while searching for words or expressions, and as a marker to introduce indirect answers. Many other studies similarly report that *well* is more frequent in the speech of English L2 learners with French, Dutch or German L1 than in L1 speech (e.g. Aijmer, 2011; Gilquin, 2008; Buysse, 2010, 2015). On the other hand, other researchers find that *well* is less frequent, rather than more frequent, in the speech of learners with Chinese or Japanese L1 (e.g., Hellermann & Vergun, 2007; Fung & Carter, 2007; Li & Xiao, 2012; Huang, 2011, 2019). A possible explanation is that the lack of a directly equivalent translation for *well* in Chinese and Japanese may lead learners to avoid using it (Liu, 2013: 157). These studies focus on *well* in adult speech. In child SLA, researchers report only low frequency of *well* in the speech of child L2 learners compared to L1 speakers (e.g. Romero-Trillo, 2002).

In the following sections, I will discuss the functions of *well* in the Ar-Children corpus compared to examples from the BNC2014 children subcorpus in light of the prior findings reviewed above. I present a quantitative analysis of *well*'s functions in both corpora in section 4.5.2, followed by a detailed look of these functions in context at the textual domain in section 4.5.3 and the interpersonal domain in section 4.5.4.

4.5.2. Frequencies of the functions of DM *well*

Table 4.7. *Raw and relative frequencies (per 100 clauses) and percentages of* well *with different functions in the two corpora*

Function		**Ar-Children**			**BNC2014 children**		
		N	**Per 100**	**%**	**N**	**Per 100**	**%**
Textual	elaborator	5	0.0	10.9	150	0.7	21.3
	filler	6	0.1	13.0	169	0.7	24.0
	transitional	9	0.1	19.6	46	0.2	6.5
	Total	**22**	**0.2**	**43.5**	**365**	**1.6**	**51.8**
Inter-	speech act	4	0.0	8.7	40	0.2	5.7
personal	responsive	10	0.1	21.7	104	0.5	14.8
	(dis)agree.	7	0.1	15.2	96	0.4	13.6
	turn management	5	0.0	10.9	87	0.4	12.4
	Total	**26**	**0.2**	**56.5**	**327**	**1.4**	**46.5**
Unclear		0	0.0	0	12	0.1	1.7
Total		**46**	**0.4**	**100**	**704**	**3.1**	**100**

As table 4.7 shows, *well* carries more interpersonal functions in the Ar-Children corpus, than in the BNC2014 children subcorpus. More specifically, *well* in the Ar-Children corpus is frequently employed as a filler, transitional, responsive or agreement/disagreement marker. In comparison, in the BNC2014 children subcorpus, it frequently functions as a filler, elaborator or responsive marker.

The differences across the two corpora between the frequencies of the functions of *well* were tested for significance using the Log-Likelihood test. To perform the test, I created a contingency table with two columns corresponding to the two corpora. In each column, I listed the frequencies of each DM subfunction, and the frequency of clauses with no DMs. According to the Log-Likelihood test, the differences between the two corpora in terms of preferences regarding *well*'s functions are significant ($p = 2.2x10^{-16}$, LL = 325.87 and df = 8). *Well* is infrequent in the Ar-Children corpus compared to the BCN2014 child subcorpus.

4.5.3. Qualitative analysis of textual functions of *well* in context

In this section and all in subsequent sections, I will apply a qualitative analysis of DMs based on the framework of DM functions discussed earlier in section 3.5.2.

As a textual marker, *well* serves three main functions in the two corpora, namely elaborator, filler and transitional functions. Elaborator *well* works as a clarification or explanation marker for a previous utterance (see section 4.5.1). This function appears only five times in the Ar-Children corpus (0.0 per 100 clauses) compared with 150 examples in the BNC2014 children subcorpus (0.7 per 100 clauses). For instance, in (1), speaker 092901 moves from talking generally about her best friends to elaborating on how she has developed her friendship with her friend's sister. In this example, *well* with rising tone contributes to the transition to providing an explanation for a previous utterance[7].

(1) 092901: mm mine I have like, two best friends, they're sisters, ***well***/ I mostly speak to the youngest/ because () I am one / I-I have, okay, so. I have a sister, and. my best friends have a sister too/ okay, and my big sister, li- uh speaks to the other big sister from my best friend. and I speak to the other one (0929/ Ar-Children)

[7] The transcription conventions used in examples from the Ar-Children corpus are given in Appendix E.

As an elaborator, *well* can also mark a justification for prior information (Schegloff & Lerner, 2009). In the Ar-Children corpus, *well* as a justification marker appears once, in the example given as (2). Speaker 084802 is shocked by her friend's confession and uses *well* to flag the words that follow as a justification for her friend ignoring her when they met in the mall.

(2) 084802: why didn't. you come up to me,

084801: cause my mum told me not to, you know, go anywhere, it was, like, a time. when we were third grade,

084802: no/ why/

084801: my mum told me not to move from my place, and I was like, okay okay. I 'm not gonna move and I saw you.

084802: no why / *oh*. okay ***well*** that's/ probably when we were, like, as best friends. right, (0848/ Ar-Children).

As noted in section 4.5.1, *well* is also used to correct or modify another speaker's contribution. Although I did not observe any instances of *well* fulfilling this function in the Ar-Children corpus, I detected 14 examples of this subfunction in the BNC2014 children subcorpus. For instance, in example (3), speakers S0655 and S0654 are collaborating to narrate the story of a book. S0655 employs *well* to mark what follows as an addition to S0654's contribution. This function of *well* overlaps with the interpersonal function of turn management (see section 4.5.4). However, I distinguish the two functions by analysing *well* as an elaborator of a previous utterance if it is preceded by another DM such as *yeah* (as here) or *okay*. In such examples, *yeah* and *okay* carry the function of turn management. My interpretation of *well* here as a non-initial marker follows Svartvik's (1980) account of *well* as a frame (see section 4.5.1).

(3) S0654: >>and then he was and then we were diving into a giant cupcake […]

S0655: yeah yeah ***well*** this this was the cupcake no this was the cupcake this is Simon () wee

S0654: wee yeah cos he missed the cupcake (S2JV/ BNC2014 children)

In the BNC2014 children subcorpus, speakers use elaborator *well* to justify not only their own decisions but also those of other people who are not involved in the ongoing conversation. In example (4), speaker S0416 employs elaborator *well* to flag a justification for a non-present friend's fondness for reading history.

(4) S0416: >>mum I doubt he was like I hate history so I 'm gonna force myself to love it I doubt he would

S0417: >>no but he probably just got got into something that other people were just it's it's very easy to do games it's not easy to get into history

S0416: fair enough ***well*** he's gifted with the talent to be good at being good at easy hard getting into history (SHX2/ BNC2014 children)

The second textual function that can be observed for *well* in the two corpora is as a filler. This function is seven times more frequent in the BNC2014 children subcorpus (0.7 per 100 clauses) than in the Ar-Children corpus (0.1 per 100 clauses). In both corpora, *well* as a filler marks self-repair or else operates as a delay technique to gain time for planning or thinking. When marking self-repair, *well* prefaces a correction or modification to a previous answer or comment (see section 4.5.1). *Well* marking self-repair appears only once in the Ar-Children corpus (example 5).

(5) 084601: did you eat breakfast.

084602: yeah. ***well.*** no, [VOC-laugh] (0846/ Ar-Children)

In the BNC 2014 children subcorpus, speakers use filler *well* as a self-repair marker in 85 examples, as in (6).

(6) S0655: although now we've discovered planet nine () I think ***well*** we have n't but there are certain signs of it (SDAB/ BNC2014 children)

Well as a self-repair marker can also mark modification with elaboration, as in (7). Schiffrin (1987: 123) calls this usage a 'background repair'.

(7) S0417: () I see if I tripped over my leg what would happen?

() S0416: I 'd go up to my room and just play on my laptop ***well*** first I 'd help help you up --UNCLEARWORD go up to my room try forget it (S8G6, BNC2014 children)

The fact that *well* is used with self-repair function only once in the Ar-Children corpus, relatively far less than in BNC2014 children, could reflect a preference on the part of the L2 children for using other DMs to flag repairs (a point further discussed in section 7.3).

Filler *well* can also indicate that the speaker is searching for words or planning upcoming discourse. In the Ar-Children corpus, speakers employ *well* as a marker of planning in five examples. This type of *well* occurs in turn-medial position with a falling tone, sometimes accompanied by other DMs. For instance, in (8) speaker 082401 uses *well* along with another DM (*actually*) as a delaying technique to think about her next question.

(8) 082401: [VOC-laugh] okay, ***well.*** actually. what is your favourite thing / [ANON-name] , you know , like , in the thing like , the larger what 's like stars , moon , trees , like that , what you love, (0824/ Ar-Children)

In the BNC2014 children subcorpus, filler *well* works as a marker of planning and appears in both turn-initial and turn-medial positions, accounting for 84 examples. Almost two fifths of the examples (38%) are accompanied by filled or unfilled pauses. For instance, *well* in (9) is accompanied by a filled pause (*erm*) and another DM (*so*).

(9) S0512: ***well*** erm so --ANONnameM erm

Example (9) is quite interesting because checking the wider context, shown in (10), reveals two other instances of *well* uttered by S0512 before the one seen in (9).

(10) S0512: --ANONnameM it was quite funny after --ANONplace

S0513 [??]: oh dear

S0512: >>***well*** yeah you d-

S0513 [??]: >>we 're not talking about that

S0512: ***well***

S0511: >>tell me about it

S0512: ***well*** erm so --ANONnameM erm

S0511: can I have some more water?

S0512: >>went down to the Trackers and there was like a bridge and (ST25/ BNC2014 children)

In example (10), S0512 is telling a story while S0513 is trying to stop S0512 from revealing its details. The first occurrence of *well* in (10) is unclear in function, due to the incomplete utterance. The speaker uses second *well* as a single turn marker to flag floor-holding (see section 4.5.4). The third *well* appears turn-initially, following S0511's request for more information, "tell me about it". By using *well*, S0512 flags the upcoming utterance as responding to that request (see section 4.5.4). Thus, *well* has interpersonal functions in these three cases. However, its being utilised in sequence with the pause and DM *so* also shows that S0512 is not completely confident about their following utterance. This use of *well* accompanied by pause in turn-initial position to flag uncertainty about an upcoming utterance is rare in the BNC2014 children subcorpus, and does not occur at all in the Ar-Children corpus. This additional function, not among those reviewed in 4.5.2, clearly

warrants further investigation. However, due to its low frequency in the data, I cannot offer any further comment here.

The last textual function of *well*, the transitional function, is not as frequent as the other textual functions in the two corpora. It appears in nine examples in the Ar-Children corpus (0.1 per 100 clauses) and 49 examples in the BNC2014 children subcorpus (0.2 per 100 clauses). In both corpora, the various uses of transitional *well* include topic changing, topic resuming, introducing direct speech and closing a turn. In my data, this kind of *well* appears with a level tone or attached to the following tone unit, with the exception of one example that carries falling tone. Speakers use this kind of *well* to fulfil their responsibility to manage conversation flow. For instance, the speaker in (11) employs *well* to indicate a transition to a new topic, talking about movies.

(11) 085301: *oh* my god/ me too. I hate vegetables. they're disgusting. [VOC-laugh]
well, do you want to go to the movies/ (0853/ Ar-Children)

Additionally, transitional *well* can signal a digression or topic resumption, as in (12). While speaker 090602 is trying to change the topic (playing Roblox) to a new one (birthdays), speaker 090601 uses *well* to resume the previous topic.

(12) 090601: I literally play Roblox all the time / and I have like this best friend in Roblox, she's like named, mm () I forgot her name. but. you can like search for it.
090602: yeah. my birthday is () I forgot. [VOC-laugh]
090601: you forgot your own birthday /
090602: mm yes, [VOC-laugh]
090601: ***well***, I love to play Roblox, and I have this (0906/ Ar-Children)

The topic change function of transitional *well* is not common in the BNC2014 children subcorpus, with only eleven examples. In some of these, it signals a shift between

aspects of the same topic, a function that is also attested by Lam (2007: 107). For instance, while discussing food ingredients in example (13), S0416 uses transitional *well* to signal a shift from talking about amounts of fats in food to amounts of sugar, a clearly related subject matter.

(13) S0417: do you know what saturates are?

S0416: erm no but it does ring a bell that name

[…]

S0416: ***well*** there's not much sugar (SHP3/ BNC2104 children)

Transitional *well* can signal a transition to quoted speech (see section 4.5.1). In the Ar-Children corpus, this function occurs only once: see (14). Speaker 084801 uses transitional *well* as a direct speech marker after quotative *be+like* to signal a transition to reporting the past speech of her friend.

(14) 084801: I do n't know. I remember when I did the hair treatment. you were , like, ***well*** my hair is more straight than yours, (0848/ Ar-Children)

Quotative transitional *well* is very rare in both corpora, with only one instance in the Ar-Children corpus and nine in the BNC2014 children subcorpus. The low rate of occurrence aligns with what the literature would lead us to expect (see Svartvik, 1980; Müller, 2004, 2005; Buysse, 2015).

Transitional *well* can also mark the closure of a topic or conversation (Carter & McCarthy, 2006: 215). In the Ar-Children corpus, speakers use transitional *well* as a closure marker in two examples, one of which is given in (15).

(15) 094501: yeah me too, ten minutes/ ten minutes, I am far. ***well***, I have- I hope that, we will uh meet each other. another time, and talk like this conversation, (0945/Ar-Children)

However, speakers in the BNC2014 children subcorpus use transitional *well* as a closure marker not only to end the conversation but also to end their own turns. In this case, *well* signals departure from the topic without initiating a new one (Heritage, 2015: 96). Example (16) shows *well* marking the end of a turn and the settlement of an argument at the same time.

(16) S0416: I only ate four

S0415: you only ate four? considering the whole packet has gone and by -- UNCLEARWORD is officially not too bad

[…] S0416: yeah probably ***well*** sorry I'm

S0415: so nonchalantly you just ate all the biscuits (S6UQ, BNC2104 children)

The results from both corpora show that children are able to use *well* with functions that assist them in achieving coherence in their discourse. Textual functions of *well* account for 46.8% of tokens of DM *well* in the Ar-Children corpus and for 51.8% in the BNC2014 children subcorpus.

4.5.4. Qualitative analysis of interpersonal functions of *well* in context

In coding the functions of *well*, I detected four interpersonal functions of *well*: speech act marker, responsive marker, turn-management function and agreement/ disagreement marker.

As a speech act marker, *well* appears before questions to indicate that the question is a request for clarification or confirmation (see section 4.5.1). *Well* signals a speech act function twice as often in the BNC2014 children subcorpus (40 instances; 0.2 per 100 clauses) as in the Ar-Children corpus (4 instances; 0.0 per 100 clauses). While it marks both clarification

and confirmation in the Ar-Children corpus, in the BNC2014 children subcorpus there is only evidence of *well* preceding confirmation questions.

In the Ar-Children corpus, *well* as a clarification marker before questions is employed to mark questions in which speakers intend to elicit more information from other speakers. This function occurs in four examples in which speakers try to maintain the flow of the conversation while requesting more details about a narrative given by their interlocutor. For instance, in example (17), while speaker 392901 is telling a story about seeing a gecko[8] in her old school, speaker 392902 interrupts with *well* to ask for details about how they got rid of the gecko.

(17) 392902: what happened to the bug,

392901: like she was uh like that/ the teacher was so mad, and then we went to the library/ and read, some books/ we came back. and then I said . I 'm never trusting this class, again/ the way she took her shoe. and like, wanted to hit the bug/

392902: ***well*** how did the bug go,

392901: 92901: it was like, I du n no / it probably / was in the floor. and then crawled/ all the way up there, which came from. (3929/Ar-Children)

In the BNC2014 children subcorpus, *well* as a speech act marker flags only confirmation questions about previously discussed information, as in (18).

(18) S0655: ***well*** did you say that's a week or a weekend?

S0653: just a weekend (SP5V/ BNC2104 children)

[8] Although speaker 392902 did not mention the word gecko in her story (using 'bug' instead), the researcher was able to infer what she was referring to from the speaker's descriptions. The use of the word bug instead of using the specific term gecko or even its correct type, i.e. insects could be a result of speaker's limited vocabulary knowledge in L2.

The second interpersonal function of *well* is the responsive function, in which *well* prefaces an answer to a question. This function is five times more frequent in the BNC2014 children subcorpus (10 instances; 0.5 per 100 clauses) than in the Ar-Children (104 instances; 0.1 per 100 clauses). By using responsive *well*, a speaker informs their hearer(s) that the forthcoming response is non-direct, non-expected or expanded (see section 4.5.1). Speakers in the Ar-Children corpus use responsive *well* to alert hearers that they are about to offer an answer with some elaboration, as in (19). By using *well*, speaker 094602 signals that the first part of the answer does not represent their complete view. Speaker 094602 subsequently uses the *but*+clause structure to indicate that the information in the latter part completes her opinion. This form appears in two examples in the Ar-Children corpus and four in the BNC2014 children subcorpus.

(19) 094601: what 's your favourite colour,

094602: ***well*** I kind of like, mm green/ but it is n't like my favourite/ but, I like it, yeah it's cool, (0946/ Ar-Children)

As a responsive marker, *well* can also indicate a speaker's lack of knowledge and flag an answer as insufficient. In (20), for instance, speaker 108101 uses *well* to signal that her answer does not align with the speaker's question.

(20) 101802: how does it, feel to be born in [ANON-place],

101801: ***well***, I really do n't remember, anything when I was there/ so I really do n't know, (1018/ Ar-Children)

Responsive marker *well* also precedes answers which might be threatening to the speaker's positive face. The question in (21) is presented as a yes/no question. Speaker 101801 prefers not to give a straightforward answer, but rather challenges the presupposition of the question. The speaker uses *well* to signal this dispreferred response (Sacks et al., 1974).

(21) 101802: that seems a really interesting one.

do you think, your parents will aglee / will agree,

101801: ***well/*** it's my room. so, it's my choice/ on how I decorate it/ (1018/ Ar-Children)

In the BNC2014 children subcorpus, speakers employ responsive *well* to preface non-straightforward and dispreferred answers – similarly to the Ar-Children examples above. Responsive *well* in the BNC2014 children subcorpus has the additional function of serving as an alert to other speakers that an answer may take more than one turn (see Heritage, 2015: 91). In this case, *well* flags the utterance it precedes as the beginning of a construction of turns to offer the answer. In (22), a speaker is given the chance to talk about a dream they have had, and prefaces their utterance with *well* to alert the other speaker that the answer requires a long narrative that may take additional turns.

(22) S0655: can I please say my dream now?

S0653: oh my goodness yes

S0655: ***well*** I'll tell you the bits I can remember it started off when this pop s- this like amazing guitar player smelt a flower and she and then but then she had to stir some sprout soup and did n't want the smell to go away […]

S0653: uhu

S0655: but then she accidentally broke loads of all her guitars () so she went on this quest

S0653: uhu

S0655: to find the guitars () and then got all these really really really weird things () hooked up like a giant cliff (S2JV/ BNC2104 children)

The third interpersonal function of *well*, as a turn management marker, also appears in turn-initial position. In this function, *well* works to alert others to the speaker's intention to

say something or to comment on another's speech. It also assists speakers in softening the effect of their comments (see section 4.5.1). As a turn management marker, *well* occurs 87 times (0.5 per 100 clauses) in the BNC2014 children subcorpus, compared with only five times (0.0 per 100 clauses) in the Ar-Children corpus. For instance, in (23), speaker 094501 uses *well* in turn-initial position to take over the conversational floor in order to describe her own relevant experience. By using *well*, she appears to be trying to comment on her friend's turn and to show her solidarity.

(23) 094502: […] uh I did n't see my grandmother / from two months, *oh* my goodness. because I was in [ANON-place] / and she 's in [ANON-place]/ uh and-and my uh mum 's dad , is in [ANON-place] , so / I did n't see the both .
094501: ***well***, I do n't see my uncle, for seven years. (0945/ Ar-Children)

In most examples in the BNC2014 children subcorpus, participants use turn management *well* to soften comments that may carry criticism or sarcasm. In (24), S0525 has been gossiping about their sister and mocking her own words. The use of *well* announces S0622's initiation of a new turn and takeover of the conversational floor. At the same time, it serves to mitigate the criticism conveyed in the other speaker's ironic utterance.

(24) S0525: mum I want your cooking () funny
S0622: ***well*** she does erm contribute a little bit towards the food (SYJK / BNC2104 children)

Additionally, speakers in the BNC2014 children subcorpus preface their opinions with turn management *well* in contexts that could be interpreted as face threatening. In this context, *well* is sometimes accompanied with other expressions, e.g. '*well* I don't know', '*well* perhaps' or '*well* maybe', and serves as a mitigator to soften the force of the speech

(see. Aijmer, 2013: 42). In (25), S0512 uses *well* to soften their opinion about their parent taking on a new job.

(25) S0511: but I applied anyway () erm and I 'm through to the final round

S0513 [??]: seems legit

S0512: >> ***well*** to be fair the job you're doing you 're getting paid more than they were wish- willing to pay for less hours than they were wanting you to () take (SASC/ BNC2104 children)

The last interpersonal function of *well* is as a marker of agreement or disagreement. In this function, *well* flags responses to other speakers' comments or narrative. This function appears in seven examples in the Ar-Children corpus (0.1 per 100 clauses) compared with 96 examples in the BNC2104 children subcorpus (0.5 per 100 clauses). When *well* signals agreement, it reinforces the other speaker's point of view. In (26), speaker 101802 uses *well* along with the DM *yeah* to express agreement regarding parental demands for them to clean their rooms. However, she follows this with the DM *but* to indicate partial agreement. Then, in the second part of her turn, she explains the difference between her experience and her friend's.

(26) 101801: you do n't get to / get screamed at / and then, they tell you. *oh* clean up / clean up your mess. none of that stuff ,

101802: ***well***/ yeah, same thing/ but the problem is I clean up my sister's stuff. [VOC-laugh] (1018/ Ar-Children)

There are six occurrences of *well* in the Ar-Children corpus that flag a disagreement. In (27), speaker 111901 employs *well* to show her disagreement with her friend's opinion. This example of *well* seems intended to serve a politeness function, as is evident if it is compared with the same utterance without *well*. Speaker 111901 is trying to soften her

disagreement by using *well* with the phrase “I think that”. Thus, disagreement marker *well* flags an upcoming objection and mitigates a potential face-threat (see section 4.5.1).

(27) 111902: how it's, like, you'll be mm sad, because you're wearing glasses. you're beautiful without glasses/

111901: ***well***, I think that I'm- that I 'm beautiful more with glasses (1119/ Ar-Children)

Agreement or disagreement marker *well* is applied differently in the BNC2014 children subcorpus. For instance, *well* as agreement marker is not often followed by other DMs such as *yes* or *yeah*. In this case, *well* indicates more than just agreement; it also indicates that the speaker has considered the options they had before announcing their agreement. For instance, in (28), *well* indicates that speaker S0655 is expressing their agreement to the offer of dinner after considering the kind of food that will be served.

(28) S0653: […] I used to mix tuna with mayonnaise and ketchup and s- finely chopped spring onions and sweetcorn and then I used to spread that mixture on the muffins […]

S0655: ***well*** then I would have that (SNPP/ BNC2104 children).

Similar to its usage in the Ar-Children corpus, 59 examples of *well* in the BNC2014 children subcorpus indicate disagreement with another speaker’s views. Example (29) illustrates *well* as a mitigating marker prior to speaker S0416’s objection to S0475’s explanation.

(29) S0475: everybody was doing yeah yeah we did well

S0416: ***well*** it was n't really like that () it was more like you you kind of played really bad you played worse than the last game (SX2V/ BNC2104 children).

All in all, analysing the interpersonal functions of *well* shows that speakers in the two corpora utilise *well* to launch a turn if they feel it may cause some confrontation with other speakers. Thus, *well* as an interpersonal marker often precedes a turn that opposes other speakers' presuppositions. This observation is clearer in the examples from the BNC2014 children subcorpus. The use of interpersonal marker *well* is rare in the Ar-Children corpus. However, the relatively few examples that could be coded as interpersonal markers show that Arab children are still in the process of acquiring *well*'s pragmatic functions.

4.6. *You know*

4.6.1. *You know* in previous literature

The DM *you know* has been studied extensively in the literature as a marker of shared knowledge (Östman, 1982; Schiffrin, 1987; Jucker &Smith, 1998; Beeching, 2016), of uncertainty (Holmes, 1986), of stalling for time to think (Erman, 1987, 2001) and as a softener (Stenström, 1990). *You know* is often associated with politeness and managing the interpersonal aspects of interactions. For instance, Östman (1982:16) argues that the main function of *you know* is a hedge that assists speakers in expressing emotions and attitudes, and indicates involvement with the ongoing conversation. According to Östman, *you know* serves face-saving functions in different turn positions. In turn-initial position, speakers employ *you know* to invoke prior agreement. When it appears with a rise or fall-rise contour in turn final position, however, it implies a tag question with a meaning such as "are you attending?", "do you agree?", or "do you see what I mean?" (Östman, 1982: 23). *You know* in this function carries rising or fall-rise intonation to communicate that the speaker seeks confirmation from the other speaker(s) (Beeching, 2016: 103). However, *you know* can alternatively appear in final position with other functions, such as flagging that the preceding

utterance contained shared information (see also Schiffrin, 1987: 276) or a "self-evident truth" (Beeching, 2016: 102).

Erman (1987, 2001) studies uses of *you know* in a corpus of teenage language and compares this to adult uses. Erman classifies these functions into three functional domains: textual, social and meta-linguistic monitor. The latter two functions are equivalent to what I refer to as the interpersonal domain. According to Erman, the textual functions of *you know* include repair, hesitation and transition between topics. Its interpersonal functions (social and meta-linguistic in Erman's terms) include its use as a marker to facilitate the negotiation of meaning between interlocutors, e.g. by turn management. It can also work as a hedge to "relieve the speaker from being completely committed to the truth value of the proposition in question" (Erman, 2001: 1341). Erman's results shows that young speakers tend to use *you know* for interpersonal functions, whereas adult speakers prefer to use it for textual functions. Beeching (2016: 109) looks at the frequency of *you know* in the speech of different age groups in the BNC1994 and in her own corpus of role play, finding that the frequency of *you know* increases with age (Beeching, 2016: 109; see Macaulay, 2002 for another similar result).

The use of *you know* varies between L1 speakers and L2 learners. Fuller (2003a: 206) points out that L2 learners employ *you know* in different contexts (e.g. narrative, dialogues) at higher rates than L1 speakers. Many other studies have also reported that L2 learners employ *you know* more than L1 speakers (e.g. Polat, 2011; Huang, 2011; Mei, 2012; Algouzi, 2014). According to these studies, L2 speakers overuse the marker *you know* to mark hesitation or planning of upcoming discourse. This indicates that although *you know* is more frequent in the speech of L2 learners, these learners have not yet mastered its various functions.

4.6.2. Frequencies of the functions of DM *you know*

Table 4.8. *Raw and relative frequencies (per 100 clauses) and percentages of* you know *with different functions in the two corpora*

Function		**Ar-Children**			**BNC2014 children**		
		N	Per 100	%	N	Per 100	%
Textual	initiator	11	0.1	12.8	7	0.0	6.1
	elaborator	25	0.2	29.1	23	0.1	20.2
	filler	15	0.1	17.4	32	0.1	28.1
	Total	**51**	**0.5**	**59.3**	**62**	**0.3**	**54.4**
Interpersonal	turn management	8	0.1	9.3	26	0.1	22.8
	shared knowledge	15	0.1	17.4	16	0.1	14.0
	speech act	12	0.1	14.0	4	0.0	3.5
	Total	**35**	**0.3**	**40.7**	**39**	**0.2**	**34.2**
Unclassified		0	0	0	13	0.1	11.4
Total		**86**	**0.8**	**100**	**114**	**0.6**	**100**

You know is most often a textual marker in both corpora, with more than half of its DM occurrences categorised as such. As a textual DM, *you know* serves mainly as an elaborator in the Ar-Children corpus. The relative frequencies of *you know* as a filler are similar between the two corpora. As an initiator, it is relatively less frequent in the BNC2014 children subcorpus. Its frequencies in the interpersonal domain – as a turn management or shared knowledge marker – are similar across the two corpora. However, *you know* as a

speech act marker is more frequent in the Ar-Children corpus than in the BNC2014 children subcorpus.

The differences across the two corpora between the frequencies of the functions of *you know* were tested for significance using the Log-Likelihood test. To perform the test, I followed the same steps as when testing *well*'s functions (see section 4.5.2). The functions of *you know* differ significantly between the two corpora at the $p<0.001$ level (LL= 41.648, df = 8).

4.6.3. Qualitative analysis of textual functions of *you know* in context

You know works as a textual marker to flag three main functions in the two corpora: initiator, elaborator and filler. The initiator *you know* is less frequent in the BNC2014 children subcorpus (0.0 per 100 clauses) than in the Ar-Children corpus (0.1 per 100 clauses). In both, initiator *you know* assists speakers in linking their new information to the ongoing discourse. *You know* often appears second, at the beginning of a turn but after another DM such as *and, but, yes* or *yeah*. This differentiates initiator *you know* from *you know* as a turn management marker (see section 4.6.4). Additionally, initiator *you know* often carries rising tone, as in (30).

(30) 090201: oh/ [VOC-laugh] okay. so.

090202: and ***you know/*** I like the presentation, sometimes when I 'm bored.

090201: okay/

090202: I go on my laptop and I do a presentation/ and tell it to my mom, [VOC-laugh] I even made a presentation for my mom's college. (0902/ Ar-Children)

The initiator *you know* often appears after other DMs in the Ar-Children corpus. Examples of *you know* in the BNC2014 children subcorpus show that the initiator function of

you know potentially overlaps with the interpersonal function of floor-holding. As such, during the coding of these examples (see section 3.5.2), I established two criteria to identify the initiator function; an example of *you know* had to fulfil both criteria to count as an initiator. First, initiator *you know* appears in turn-initial position and introduces new information not previously mentioned in the discussion. Second, initiator *you know* appears in utterances produced by speakers who have been mainly listeners and have not actively contributed to the ongoing discussion before their actual participation (see section 4.6.4 for comparison with other interpersonal functions of *you know*). By these criteria, seven instances of *you know* in the BNC2014 children subcorpus functioned as initiators.

In (31), while speaker S0417 was listing their favourite school subjects and stating the reasons for that, speaker S0416 was mainly a listener. *You know* flags speaker's S0416 introduction of the new information about sports.

(31) S0417: to pick (...) and I did sports but I was n't really into them but I was good at them and I was always being made a sports captain of my form and
S0416: that's good
S0417: I was kind of lucky actually to be honest looking back I was lucky I was kind of naturally okay at most things at school and I really liked it
S0416: ***you know*** most of the girls? they they hate sports
S0417: mm
S0416: like they really hate the like --ANONnameF got a note to go out of PE
(SQNA / BNC2014 children)

The second textual function of *you know* is as an elaborator. We find 25 instances of elaborator *you know* in the Ar-Children corpus (0.2 per 100 clauses) compared to 23 instances (0.1 per 100 clauses) in the BNC2014 children subcorpus. This function appears in

turn-medial position to flag a speaker's attempt to elaborate on their prior speech through explanation, as for instance in (32), or exemplification, as for instance in (33). In (32), Speaker 082401 uses *you know* prior to explaining her previous turn to ensure that her friend has grasped the grade of the colour she has been describing.

(32) 082401: purple with white, ***you know***, you mix it with white/ I like that colour,

082402: See, purple with white, (0824/ Ar-Children)

In (33), speaker 120801 employs *you know* before elaborating on her previous turn by giving examples of places to which she has travelled. *You know* is employed to strengthen the impact of her examples and encourage the other speaker to appreciate the extent of her achievement.

(33) 120801: oh travel / I travel too much,

120802: try- I traveling uh () I travel in, I dunno. I do n't travel anywhere, in airplane,

120801: I travel so many times. ***you know*** one time I went to a six-hour , drive/

(1208/ Ar-Children)

Similar to its usage in the Ar-Children corpus, in the BNC2014 children subcorpus elaborator *you know* marks a speaker's attempt to explain or exemplify a previous utterance, as in (34), where Speaker S0622 exemplifies the dangers they may face on their trip.

(34) S0622: >>then *well* then you're then they would n't go over the bridge er obviously for safety reasons ***you know*** if the bridge collapsed with the the

S0641: >>yeah

S0622: weight of the train they'd have to er () get it checked out or mend the bridge and that could take a day or two at least () (SDZC/ BNC2014 children)

You know as a filler works to signal a self-repair or planning. The frequency of filler *you know* is similar in the two corpora (0.1 per 100 clauses), with 15 instances in the Ar-Children corpus and 32 in the BNC2014 children subcorpus. In both, the planning subfunction appears more frequently than the self-repair subfunction. As a self-repair marker, filler *you know* expresses a speaker's desire to revise and adjust their prior utterance. In (35), for instance, speaker 082401 uses *you know* to mark a modification to the subject of the restarted clause. This subfunction is infrequent in the Ar-Children corpus (just five examples).

(35) 082401: yeah, some people just have, ***you know***, some families just have one-one children / (0824/ Ar-Children).

Filler *you know* with the planning subfunction, where its use stalls for time to think of an appropriate word or phrase, likewise occurs in the Ar-Children corpus. In this subfunction, *you know* co-occurs with filled and unfilled pauses, or with other DMs. In (36), speaker 082101 employs *you know* with preceding and subsequent filled pauses, signalling her attempt to think about what she wants to say next.

(36) 082102: yes / I love Greg.
082101: Rowley Rowley.
082102: [VOC-laugh] he's funny,
082101: yeah, mm ***you know***/ uh, I read the story that, uh, [ANON-name] told us. (0821/ Ar-Children)

There are 11 and 21 instances respectively of the self-repair and planning subfunctions of filler *you know* in the BNC2014 children subcorpus. In (37), *you know* is used between two filled pauses while the speaker takes time to think about the explanation they will give.

(37) S0655: they might have been able to cos the most er erm nearly crack the code on that th- or on the thing that –ANONnameF

S0654: what code?

S0655: erm ***you know*** erm get those erm things where you can line up different letters like A is T (SADE / BNC2014 children)

The textual functions of *you know* appear in more than half of its instances as a DM in the two corpora. This result adds to our understanding of how children use DMs to fulfil functions of coherence in their communication.

4.6.4. Qualitative analysis of interpersonal functions of *you know* in context

As an interpersonal marker, *you know* works as a turn management marker, shared knowledge marker and speech act marker. Turn management *you know* appears with similar relative frequency in the two corpora (0.1 per 100 clauses). When it has this function, *you know* appears in turn-initial position to express the speaker's desire to take over the turn or to return to completing their previous turn. By employing *you know* for turn management, speakers indicate that they have something to share with other speakers. For instance, in example (38), while speaker 125701 is describing her skills in gymnastics, the other speaker inserts a comment on her sister's skills. Then, *you know* expresses speaker 125701's desire to re-take the conversational floor to describe her experience in learning the "splits".

(38) 125701: and I'm learning the cartwheel/ too,

125702: yeah, my- my sister knows how to, like do this cartwheel, and splits.

125701: ***you know***/ always/ I try to practice on the split/ and my- like, my legs are almost close, they're so close/ to just like open, (1257/ Ar-Children)

In the BNC2014 children subcorpus, speakers use turn management *you know* to elaborate on previous turns and to demonstrate collaboration in the ongoing discussion by adding clarification, justification or more details to prior turns, as in (39).

(39) S0416: but I do n't believe in Scientology cos Scientology is crazy

S0417: >>science is nothing to do with Scientology it's just their names are similar

S0416: >> ***you know*** Scientology believes that there was an alien mothership called Glorgon that like had a whole thing created the Earth and all that stuff (S8Q3/ BNC2014 children)

Unlike the turn management function, *you know* marking shared knowledge and speech act function often appears in turn-final position. As a marker of shared knowledge, *you know* flags speakers' attempts to establish common ground with the other speakers and express solidarity. The frequencies of *you know* as a marker of shared knowledge are similar in the two corpora (0.1 per 100 clauses). In this function, speakers do not prompt a verbal response to come after *you know*. In (40), for instance, Speaker 084601 has been telling her friend about a time when she was crying because she witnessed her little sister fall from the table and consequently injure her head. She is explaining to her friend the reason for her fear at that time. In this example, speaker 084601 uses *you know* with a falling tone shows that she is unconfident and reluctant to say the word aloud, (e.g., because it is such an awful thing) and thus uses *you know* to encourage her friend to infer the meaning.

(40) 084602: why/[VOC-laugh]

084601: because I thought she was. ***you know***.

084602: dying,

084601: but then, I blamed the maid, and its actually her fault, because she is supposed to watch my sister. (0846/ Ar-Children)

Shared knowledge marker *you know* can sometimes appear in turn-medial position. This often occurs with falling or level tone. For instance, in (41) the speaker has been struggling to describe the situation of her family as "separated" in English. She first uses *you know* as a filler while she attempts to search for the word (see section 4.6.3). Then she employs *you know* a second time to alert her friend that the thing for which she does not know the right English word is something known to both of them, and supplies the Arabic equivalent موزعة *moazaa* 'separated'.

(41) 103702: yeah, it's very nice/ my-my family, my family's like, this, () you know, like, I do n't know it's in English, but, ***you know.*** it's like, موزعة/ (1037/ Ar-Children)

Coding this function necessitates consulting its prosody but prosodic annotation is not available for the BNC2014 children subcorpus. Therefore, I decided to differentiate between this function and *you know* as a speech act marker by one defining rule: I coded any occurrence of *you know* as a shared knowledge marker if it appears in turn-final position and is not followed by an affirmative or negative answer from another speaker. I thus coded 16 instances of *you know* in turn-final position as shared knowledge markers in the BNC2014 children subcorpus, including (42).

(42) S0655: I like five second rule and I 'm glad it 's not one hour rule because then it would take quite a while for some food to ***you know***
S0653: >>I have I've got you some water it 's right next to you (SCP7/ BNC2014 children)

You know as a speech act marker has been analysed in previous studies as functioning like a tag question (see section 4.6.1). *You know* in this function carries a rising or fall-rise intonation contour. There are 12 such examples in the Ar-Children corpus. Speech act marker *you know* does not necessitate a response but does indicate an attempt to engage other speakers in the discourse. In (43), speaker 094602 is aware that her friend has never travelled

outside the country. She tries to engage her friend in the conversation by using *you know* at the end of her turn as a confirmation speech act marker.

(43) 094601: it's my dream to , go to uh Disneyland, but like

094602: it's really good. they have hotels there, they 're really nice , and everything is cool ***you know/***

094601: in London/

094602: yeah. and also in Paris (0946/ Ar-Children)

The lack of prosodic information in the BNC2014 children subcorpus makes it difficult to code this function. I only coded *you know* in turn-final position as a speech act marker in the BNC2014 children subcorpus if the speaker got an affirmative response, illustrating clear use of *you know* with the effect of a tag question, as in (44). But there were only four examples like this (0.0 per 100 clauses).

(44) S0416: it's like white and it has like black tessellation do *you know* that one? and it's and it has a black pocket ***you know***?

S0417: oh yeah (SDFR / BNC2014 children)

In sum, in both corpora *you know* works as an interpersonal DM to assist children in engaging the hearers in the conversation. Interpersonal functions account for less than half of the instances in both corpora (40.7% in the Ar-Children corpus and 34.2% in the BNC2014 children subcorpus). This result shows that the children represented in these corpora are aware of *you know*'s functions as a textual and interpersonal marker.

4.7. *I mean*

4.7.1. *I mean* in previous literature

I mean is described in the literature as mainly a textual DM – one used to signal a speaker's modification of their own talk (Schiffrin,1987: 296). According to Schiffrin (1987),

this can take the form of elaboration of their own speech or repair of a prior utterance. Brinton (2007) observes other discourse functions of *I mean* at the textual level: as a marker of exemplification, and as a cause marker paraphrasable as "I am saying this because". Moreover, *I mean* can be used to emphasise the truthfulness of an utterance.

Erman (1987: 145) differentiates the various functions of *I mean* according to its position. *I mean* has functions at turn level or "micro level" as well as at discourse level or "macro-level". At both levels, *I mean* can serve as a clarification or mitigation marker for previous utterances or discourse. At a discourse level, *I mean* signals a justification of a previous claim. But its functions at turn level differ depending on its position within the turn. In turn-initial position, *I mean* works as a turn-taking or topic-shifting marker, whereas in turn-medial position, *I mean* refers to upcoming modification for prior arguments. *I mean* in turn final position is rare, and is mainly a floor yielding technique (Erman, 1987: 54). In contrast, Forschini (2010) finds that *I mean* in turn-initial position is more than a turn management marker – it can introduce a clarification of a prior turn.

4.7.2. Frequencies of the functions of DM *I mean*

Table 4.9. *Raw and relative frequencies (per 100 clauses) and percentages of* I mean *with different functions in the two corpora*

Function		**Ar-Children**			**BNC2014 children**		
		N	Per 100	%	N	Per 100	%
Textual	elaborator	15	0.1	60.0	60	0.3	51.3
	Filler	4	0.0	16.0	17	0.1	14.5
	Total	**19**	**0.2**	**76.0**	**77**	**0.3**	**65.8**
Interpersonal	turn management	6	0.1	24.0	22	0.1	18.8
	Total	**6**	**0.1**	**24.0**	**22**	**0.1**	**18.8**
Unclassified		0	0.0	0	18	0.1	15.4
Total		**25**	**0.3**	**100**	**114**	**0.4**	**100**

In both corpora, *I mean* is more often a textual DM, albeit both textual functions are more frequent in the BNC2014 children subcorpus than in the Ar-Children corpus. The only interpersonal function of *I mean* observed in either corpus is turn management.

The differences across the two corpora between the frequencies of the functions of *I mean* were tested for significance using the Log-Likelihood test. According to the Log-Likelihood test, there is no significance difference between the frequencies of the functions of *I mean* in the two corpora at the p <0.001 level (p = 0.00343, LL= 22.974 and df = 8).

4.7.3. Qualitative analysis of textual functions of *I mean* in context

As a textual marker, *I mean* serves elaborator and filler functions in both corpora. Elaborator *I mean* flags intent to clarify, reformulate or justify prior utterances. This is the most prevalent function of *I mean* in the two corpora. However, it appears three times more frequently in the BNC2014 children subcorpus (0.3 per 100 clauses) than in the Ar-Children corpus (0.1 per 100 clauses). Speakers employ elaborator *I mean* turn-medially mainly to clarify or explain preceding part of the same utterance. For instance, in (45), speaker 091802 uses *I mean* to clarify the question that she asked her friend about, i.e. her experience with travel.

(45) 091802: where have you travelled, ***I mean*** like not holidays, travel/ (0918, Ar-Children)

Elaborator *I mean* can also appear in medial position to mark an attempt to justify, rather than clarify, a prior utterance. Speaker 085301 uses *I mean* before her justification of why she hates school. Thus, *I mean* introduces a justification of the prior answer.

(46) 085302: do you love school/

085301: not that much. ***I mean***, getting up at 6:00 AM. kind of sucks. (0853/ Ar-Children)

Elaborator *I mean* can also flag an upcoming reformulation of a previous part of the same utterance. In (47), speaker 084802 describes to her friend a place they visited together, but uses the word “windows” inappropriately. The use of *I mean* flags that she intends to reformulate her description, which she then does.

(47) 084802: and it has windows, that not windows. ***I mean*** mirrors all over the place.

084801: they are like. funny/ (0848/ Ar-Children)

In the BNC2014 children subcorpus, elaborator *I mean* occurs in 40 instances. What is interesting is that some of these instances appear in turn-final position and work retrospectively as an elaborator. Example (48) shows the use of *I mean* in turn-final position to mark a speaker's attempt to clarify what they are thinking. Fox Tree (2002: 742) describes *I mean* in turn-final position as an attempt of speakers to explain they mean something else, but "leave the adjustment off record". In this example, the repetition of related terms (i.e. two words ending with "-ism", repetition of "gay", and the two versions of "illegal") might be interpreted as S0512 avoiding mention of the topic at hand and using only vague references: "stuff like that". However, this cannot be the case, since they readily affirm and repeat S0511's suggestion of an appropriate word ("homophobia"). Instead, S0512 may have been searching for the right single word to summarise the ideas they are referring to, and thus may have employed *I mean* in turn-final position to flag this attempt at clarification.

(48) S0512: yeah when I think of Russia

S0511: yeah

S0512: I think of racism () er gay illegalism () gay --UNCLEARWORD illegal and stuff like that ***I mean***

S0511: homophobia's the word you're after

S0512: >>yeah homophobia there you go (SP3F/ BNC2014 children)

The other textual function of *I mean* in both corpora is its use as a filler, indicating hesitation, repair or planning. Filler *I mean* is more frequent in the BNC2014 children subcorpus (0.1 per 100 clauses) than in the Ar-Children corpus (0.0 per 100 clauses). Four instances of *I mean* in the Ar-Children corpus have the function of attempting to gain time to plan their speech. With this function, filler *I mean* can appear with preceding or following filled pauses, or with both, as in (49).

(49) 122402: so what's your favourite school subject.

122401: school subject/ my favourite- favourite school subject/ is uh math/ uh ***I mean***, uh Islamic/ and science. okay, uh which- uh which movie do you like. (1224/ Ar-Children)

It also can appear with a full repetition of a prior utterance as in (50).

(50) 110502: do you have- ***I mean*** do you have friends /

110501: yeah, () do you have pets/ (1105/ Ar-Children)

In the BNC2014 children subcorpus, filler *I mean* co-occurs with filled pauses and/or with utterance repetition in 17 instances to express hesitation, planning or to repair syntactic errors. In (51), *I mean* marks hesitation alongside repetition of DM *like* and preposition *for*.

(51) S0416: right it is kind of like super like like weird but I think like I love I totally take advantage of like coming and doing sports because like ***I mean*** like for for you you 're getting like (SQNA/ BNC2014 children)

I mean can also mark a speaker's attempt to repair a preceding part of the same utterance. For instance, in (52), *I mean* marks the speaker's amendment of "they'd" to "they were". This speaker is aware that a repair is needed and they go on record by introducing it with *I mean*.

(52) S0622: they'd erm ***I mean*** they were only there as a er a guide and a bit of offer a bit of () erm help (SYJK / BNC2014 children)

4.7.4. Qualitative analysis of interpersonal functions of *I mean* in context

In contrast to findings of previous studies, the only interpersonal function of *I mean* detected in my two corpora is the turn management function (see section 4.7.2). The relative frequencies of this function are similar in both (0.1 per 100 clauses). *I mean* with turn

management function occurs in turn-initial position to mark the organisation of turn-taking between interlocutors. By using *I mean,* speakers display their desire to take over the conversational floor to offer a clarification or justification. In the Ar-Children corpus, 6 instances of turn management marker *I mean* appear in turn-initial position. Example (53) illustrates different variants of *I mean*'s turn management function. The first instance of *I mean* signals speaker 101802's attempt to retake the conversational floor, to offer a clarification on her prior turn. In her next turn, speaker 101802 employs *I mean* a second time to successfully take the conversational floor, to elaborate on her argument and justify her negative opinion about AirPods.

(53) 101802: mm not that much, we do n't use devices really. I only, have like my own headphones, and my own, () phone, same with my sister. but she have AirPods,

101801: that's cool,

101802: ***I mean***. I do have one, my mother gave it to me. but I do n't like using it that much, for some reason, () they would always fall,

101801: I my mum / has AirPods / so I always, take them from her. and I just use them , […]

101802: ***I mean***, yeah, it's kind of fun to have AirPods, because, they do n't bother you/ […] I use it/ but like, most of the time, they would fall off my ears, so my mother bought me a new one that never fall, () (1018/ Ar-Children)

Turn management *I mean* also marks speakers' comments on the arguments of others. For instance, in (54), speaker 080201 complains about being an eldest sister and being annoyed by her family members. Speaker 080202 then employs *I mean* twice to take over the conversational floor and offer her point of view. *I mean* serves as a mitigator for the force of the upcoming utterance as a face threatening act as well as having a turn-management

function. The speaker immediately tries to offer an opposing comment and to be cooperative in the conversation.

(54) 080202: exactly/ they- it's, what's siblings do when they're bored,

080201: like , I try to be , I I know what they 're doing when they come to annoy me , they do n't just see me mad , they want me to hit them.

[…] 080202: okay, ***I mean***/

080201: but of course I love my siblings.

080202: ***I mean***, they're your family. (0802/ Ar-Children)

Similar to its usage in the Ar-Children corpus, in the BNC2014 children subcorpus turn management *I mean* expresses a speaker's desire to take over the conversational floor in order to offer a clarification for their prior turn, for example, as in (55).

(55) S0653: well if you could eat something that was just a number I would be very impressed

S0655: just a number

S0653: mm

S0655: ***I mean*** is it that years old? (S9UA / BNC2014 children).

Additionally, in the BNC2014 children subcorpus turn management *I mean* can mark a polite comment or opinion. For instance, in (56), speaker S0416 uses the first instance of *I mean* to interrupt their parent's turn and offer their point of view. The second *I mean* flags the speaker's attempt to take over the conversational floor and state explicitly that the prior utterance was their own opinion.

(56) S0415: and then some --UNCLEARWORD took over it and everyone said this is ridiculous we need to revamp this up and sort of get people to come back and come

S0417: >>yeah

S0415: and give –UNCLEARWORD

S0416: >>***I mean*** a terabyte is pretty good

S0415: that's ridiculously good

S0416: ***I mean*** my opinion (S6UQ / BNC2014 children)

Turn management is the only interpersonal function evident for *I mean* in the two corpora. But due to the low frequency of *I mean*, this does not constitute evidence of whether or not the children, L1 or L2, have yet acquired the various strategies for use of *I mean* as a turn management marker. It is possible that they use other markers to fulfil this function. I will discuss this issue further in section 7.3.

4.8. Summary

In this chapter, I have provided an overview of my data on the frequencies and textual/interpersonal functions of DMs in the Ar-Children corpus and the BNC2014 children subcorpus. First, I answered my RQ1 concerning the most frequent DMs in the two corpora. The main findings were that (1) *like* and *yeah* are the most frequent forms that can be DMs in the two corpora; and (2) comparison of relative frequencies shows that *like* and *okay* are the most frequent DMs in the Ar-Children corpus, compared to *oh* and *yeah* in the BNC2014 children subcorpus. A similarity between the two corpora is the low frequency of *you know* and *I mean*.

In section 4.3, I discussed the frequency of these DMs in the speech of Arab children in relation to the context of their studying English in preschool (in English-speaking versus Arabic-speaking countries) and number of years of preschool in English-speaking countries. This part of the discussion answers RQ2. The results show that participants who attended English-medium preschool in Arabic-speaking countries show a higher frequency but less

variety in their use of the eight DMs. This result is consistent in the two different age groups (9-10 and 11-12 year olds) in the Ar-Children corpus.

In section 4.4, I analysed the use of each DM according to the textual and interpersonal functional domains. This section partly answers RQ3. I found that four out of the eight DMs carry textual functions in both corpora, namely *like, so, you know* and *I mean*. *Yeah*, *okay* and *oh* are mainly interpersonal DMs in both corpora. Meanwhile, the functions of *well* differ between corpora: mainly interpersonal in the Ar-Children corpus; mainly textual in the BNC2014 children subcorpus. However, this finding regarding *well* should not be interpreted as conclusive due to the low frequency of *well* as a DM in the Ar-Children corpus.

After this quantitative analysis, I moved on in sections 4.5, 4.6 and 4.7 to a qualitative analysis of the less frequent DMs, namely *well, you know* and *I mean*. This analysis likewise offers part of the answer to RQ3. The analysis of *well* shows that it is more frequent as a responsive and transitional marker in the Ar-Children corpus and as a filler and elaborator in the BNC2014 children subcorpus. *You know* is frequently used as an elaborator in the Ar-Children corpus. Lack of prosodic information in the BNC2014 children subcorpus complicated the process of distinguishing between some of the functions of *you know*, i.e. shared knowledge marker versus speech act marker. Lastly, while in the textual domain *I mean* was seen to carry different functions, in the interpersonal domain, we observed the turn management function only.

The analysis of *well*, *you know* and *I mean* shows that L1 and L2 child speakers use these three markers with similar functions. However, there are examples of some rare subfunctions in each corpus. In addition, there is evidence of overlap between the two functional domains, i.e. textual and interpersonal. I postpone further discussion of these

findings until I have introduced my findings regarding the other five DMs, so that the discussion (in chapter 7) can consider and compare all these results together. Over the next two chapters, I present analyses of the more frequent DMs *like, so, yeah, okay* and *oh*, with illustrative examples from both corpora. These analyses complete the response to RQ3.

Chapter 5: *Like* and *So*

5.1. Overview

This chapter discusses the functions of DMs *like* and *so*, as the most frequent DMs in the Ar-Children corpus. The discussion of *like* and *so* functions offers part of the answer to the third RQ: *what are the functions of the most frequent DMs in the speech of these Arabic-speaking school-age children? Are there any differences in the functions with which DMs are used, in comparison to native English-speaking children?* These two markers are discussed together in the present chapter because their functions are mostly textual in the Ar-Children corpus (see section 4.4). Chapter 6 will discuss the remaining DMs, *yeah*, *okay* and *oh*, which largely perform interpersonal functions in the Ar-Children corpus.

Sections 5.2 and 5.3 present *like* and *so* respectively. Similar to the previously discussed DMs *well*, *you know* and *I mean* (sections 4.5, 4.6 and 4.7 respectively), each main section is structured as follows:

1. First, I present a review of previous literature concerning the DM under discussion.
2. Then, I discuss the frequencies of the DM's functions and subfunctions in the two corpora.
3. Finally, I give a detailed qualitative analysis of the DM's functions with examples from the two corpora.

In section 5.4, I summarise the results for *like* and *so*, pointing to the major differences in how each DM is used between the two corpora.

5.2. *Like*

5.2.1. *Like* in previous literature

The DM *like* is highly frequent in spoken English and, consequently, one of the most extensively studied DMs (Diskin-Holdaway, 2021: 563). Overuse of DM *like* in speech has been perceived negatively as a marker of young or uneducated speakers, or as a marker of bad English (Dailey-O'Cain, 2000: 75). As a DM, *like* has also been described as hesitational or a "garbage" word (Fox Tree, 2006: 726). However, researchers have described many functions of *like* in discourse (e.g. Underhill, 1988; Schouroup, 1985; Miller & Weinert, 1995; Jucker & Smith, 1998; Miller, 2009; Ranger, 2018). Many studies classify *like* as mainly a textual DM that works only at discourse level (e.g. Müller, 2005; Algouzi, 2014; Özturk, 2018). However, others claim that *like* can express interpersonal functions, as a device to establish intimacy between speakers (e.g., Siegel, 2002, Miller, 2009: Andersen, 1998, 2001; Beeching, 2016; D'Arcy, 2017).

The DM *like* serves many different functions. It can act as a focuser, a filler, and a marker to introduce examples or explanation. As a focuser, *like* either introduces new or unusual information (e.g. "we went over to this house to get some ***like*** *plywood*", example 6f in Underhill, 1988), or marks a focus on the significant piece of information (e.g. "I couldn't make it to the class 'cause, *well, **like***. I had this accident on the freeway", example 51 in Underhill, 1988) (see also Romaine & Lange, 1991; Dailey-O'Cain, 2000; Miller & Weinert, 1995). The focuser function was first described by Underhill (1988), who observes that *like* can be used as a marker of newness. According to his analysis, focuser *like* is often located non-initially in the utterance. Using *like*, speakers can distance themselves from unusual or technical terms that are introduced as new information. For instance, in the example "Either I'll hafta replace the arm or I'll hafta ***like*** uh realign the arm", *like* precedes the verb *realign*,

which is a more technical concept than the preceding alternative *replace* (Underhill, 1988: 237). As a focuser, *like* can also mark exaggeration that is the utterance should not be taken seriously (Underhill, 1988: 242; see also Fuller, 2003c; Müller, 2005, Beeching, 2016), as for instance in "I'm really going to rest this weekend. I mean ***like*** stay in bed all day Saturday and Sunday" (Underhill, 1988: 241).

As a textual marker, *like* can also be a filler or a marker to introduce examples or explanation. As a filler, *like* has functions related to planning difficulties (e.g. "Alright, erm, *well* ***like***, I usually take the train about… twenty past", example 183 in Andersen, 2001: 270), such as repair (e.g. "and he listed ***like*** reeled off a load of blokes"), false start (e.g. "I used to ***like*** we used to see each other", Andersen, 2001: 270) or search for appropriate word or phrase (Andersen 2001: 255). As a marker of exemplification, *like* works as a substitute for the phrase *for example* in spoken conversation (e.g. "if there's a fight, they come back with blades and that and then ***like,*** baseball bats, hammers, and they get ready for a fight", example 135 in Andersen, 2001: 236) (see also Schouroup, 1985: 36; see also Miller, 2009: 336; Andersen, 2001: 236; Fung & Carter, 2007: 424; Ranger, 2018: 253). This function is explained from a relevance theoretical view (see 2.3.2.2) by Jucker & Smith (1998). In their account, *like* signals that the following utterance should not be interpreted literally, but rather as a set of hints that assist listeners in constructing their interpretation (Jucker & Smith, 1998: 185, see also Andersen, 2001: 236). As an explanation marker, *like* flags the upcoming utterance as an extension of their prior speech (Müller, 2005: 215). In this function, speakers use *like* either to *reformulate* their speech (e.g. "a shot at the dining room ... *you know* ***like*** this random shot", example 200 in Müller, 2005: 216), to explain it by providing an *alternative term* to one given previously (e.g. " the.. artist takes his change and leaves a big coin ***like*** a silver coin", example 202 in Müller, 2005: 215), or to *repeat* the previous

utterance with almost the same wording (e.g. “what was your favourite part of it? ***like*** the funniest part”, example 204 in Müller, 2005: 238).

Researchers have examined the different functions of *like* according to its position. For instance, Schouroup (1985: 40) argues that filler *like* (or *pausal interjectional* in Schourup’s terms) flags planning difficulties in initial as well as medial positions. Ranger (2018: 253) argues that *like* as a marker of examples has two core functions according to its position: an *argumentative* and a *meta-representational* function. Argumentative *like* appears in utterance peripheral position, and connects the host utterance with a previous one to provide examples of a more general phenomenon (e.g. “Andy knows everything! ***Like***, Andy knows about me”, example 32 in Ranger, 2018: 245). Meta-representational *like* occurs clause-medially as a focuser (e.g. “They think America is ***like*** Disney World”, example 34 in Ranger, 2018: 246). Another view of *like* functions according to their positions is introduced in Miller & Weinert (1995: 390). They argue that *like* in initial position marks speakers’ attempts to offer explanation in declaratives or seek more information in interrogatives, whereas clause-final *like* carries the role of “clearing up misunderstanding” or confronting false assumptions or oppositions (Miller & Weinert, 1995: 392). In this way, we can conclude that *like* can work as an interpersonal marker, facilitating communication between interlocutors.

Much research on DM *like* has addressed whether it is used differently by adults versus adolescents. Andersen (2001: 247) argues that young speakers’ preference for marking new information with *like* is to express their uncertainty about that information and its appropriateness. Miller (2009) finds no difference between the use of *like* in adult and adolescent speech; rather, both groups use *like* to highlight new information and “give it additional rhetorical and dramatic force” (Miller, 2009: 334). Siegel (2002) argues that focuser *like* can also serve at interpersonal level as an intimacy marker between young

speakers. Miller (2009: 334), however, criticises Siegel's view on the basis that his data came from his teenage daughter and her friends, whereas Miller observes this function on conversations between strangers on phone calls on the radio who comprises two different age groups: adults and adolescents.

Like as a DM has also been of interest to SLA researchers (e.g. Algouzi, 2014; Özturk, 2018). These studies report that the various functions of *like* are more frequent in the speech of L1 speakers than that of language learners. By contrast, Fung & Carter (2007) report that the frequency of *like* is comparable in their corpus of Chinese learners and in a corpus of British English speakers (a subcorpus constructed from CANCODE). Likewise, Diskin (2017) reports that *like* occurs with comparable frequency in the English speech of Polish and Chinese migrants in Ireland and in the speech of Irish L1 speakers of English. However, they further report that the L2 group use *like* with high frequency in initial and medial position, but not final position. Use of *like* in final position is recognised as a feature of Irish English that their L2 group has not fully acquired. Diskin (2017:155) argues that in order to acquire DM *like* in final position, a speaker must first be aware of its pragmatic function as a *mitigating* device. According to Diskin, *like* serves as a mitigator in clause final position to soften the force of an utterance or to avoid threatening the listener's positive face. This implies that *like* in final position can be an interpersonal marker, as earlier suggested by Miller & Weinert (1995) this interpersonal function was not evident in the speech of L2 speakers studied by Diskin.

In summary, according to previous studies, *like* is mainly a textual marker, but some researchers argue that it also has interpersonal functions. From the fact that different authors analyse clause-final DM *like* differently, we can conclude that the distinction between textual *like* and interpersonal *like* is not clear cut. In the present study, I opt to follow the prevailing research and analyse *like* primarily as a textual marker, while highlighting examples that may also have interpersonal function.

5.2.2. Frequencies of the functions of DM *like*

The frequency of *like* as DM and non-DM is presented in table 5.1. Non-DM use of *like* has been discussed in section 3.5.1; in this section, I focus on *like* as a DM.

Table 5.1. *Raw & relative frequencies (per 100 clauses) of* like *in the Ar-Children corpus and the BNC2014 children subcorpus*

Corpus	Total use of *like*	DM use	Non-DM use
Ar-Children	2,088 (26.5)	1,199 (11.1)	889 (15.4)
BNC2014 children	2,306 (17.2)	1,251 (5.2)	1,055 (12.0)

Table 5.1 shows that participants in the Ar-Children corpus use *like* more frequently as a DM than do participants in the BNC2014 children subcorpus. Table 5.2 gives the raw and relative frequencies and percentages of *like*'s functions in each corpus.

Table 5.2. *Raw and relative (per100 clauses) frequencies and percentages of* like *with different functions in the two corpora*

Function		**Ar-Children**			**BNC2014 children**		
		N	Per 100	%	N	Per 100	%
Initiators		**418**	**3.9**	**34.9**	**383**	**1.7**	**30.6**
Elaborators	exemplifier	70	0.7	5.8	154	0.7	12.3
	explanator	377	3.5	31.4	476	2.1	38.0
	Total	**359**	**3.3**	**29.9**	**498**	**1.8**	**39.8**
Fillers	search	271	2.5	22.6	242	1.1	19.3
	false start	128	1.2	10.7	92	0.4	7.4
	Total	**399**	**3.7**	**33.3**	**334**	**1.5**	**26.7**
Unclassified		23	0.2	1.9	36	0.2	2.9
Total		**1,199**	**11.1**	**100**	**1,251**	**5.2**	**100**

The data indicate that *like* largely fulfils textual functions, although some examples with interpersonal functions were observed. This is in line with the prior literature discussed in 5.2.1. Speakers in the Ar-Children corpus employ the textual functions of *like* – initiator, elaborator and filler – three times more often than speakers in the BNC2014 children subcorpus. The largest proportions of *like* in the Ar-Children corpus were coded as initiators, followed by fillers, and then elaborators. In the BNC2014 children subcorpus, on the other hand, the order is elaborators, followed by initiators, and then fillers.

The differences across the two corpora between the frequencies of the functions of *like* were tested for significance using the Log-Likelihood test. I created a contingency table with two columns corresponding to the two corpora. In each column, I listed the frequencies

of *like* occurring with each subfunction, plus the frequency of clauses with no DMs. According to the Log-Likelihood test, the differences between the two corpora in the functions of *like* were highly significant at the $p<0.001$ level ($p = 2.2x10^{-16}$, LL= 372.43, df = 7).

5.2.3. Qualitative analysis of textual functions of *like* in context

In this section, I discuss the use of DM *like* with the functions of initiator, elaborator, and filler (see section 3.5.2), including sub-categories.

5.2.3.1. Initiator *like*

In both corpora, *like* is used with initiator function, that is, to introduce new information. This function appears more frequently in the Ar-Children corpus than in the BNC2014 children subcorpus; an example from the Ar-Children corpus is given as (1), in which the speaker uses *like* to mark "the mall" as new information.

(1) 084601: […] and then my auntie and my cousins came to pick me up, we went to ***like*** a mall, to cheer me up, (0846/ Ar-Children).

In (2), we see an example of the same function in the BNC2014 children subcorpus. Speaker S0416 uses *like* each time they introduce a new item of information about a fun place they have been to.

(2) S0417: I have n't been there for ages

S0416: and then it 's like there's a green screen and you go up there and the green screen and you can get ***like*** one of those cameras and you can use that as a hand and there's ***like*** a big control panel where you can change what 's on the green screen (S3CK / BNC2014 children)

5.2.3.2. Elaborator *like*

Elaborator *like* is more frequent in the Ar-Children corpus (3.3 per 100 clauses) than in the BNC2014 children subcorpus (1.8 per 100 clauses). Speakers use elaborator *like* prior to elaborate on previous utterances with examples, provide further explanation, or focus on previous information. These functions are termed *exemplifier*, *explanator* and *focuser*, respectively as previously explained in section 5.2.1.

The exemplifier function of *like* is shown in (3), where the speaker elaborates on her question by giving examples of possible acceptable answers.

(3) 101501: okay, uh uh what grade did you like the most, ***like*** grade one, grade two three five six/ (0941/ Ar-Children).

The exemplifier function can also involve illustration, or giving an example that does not expand on a previously mentioned concept or a definite list (Müller, 2005: 213; see also Beeching, 2016: 129 discussed earlier in section 5.2.1). For instance, in (4), speaker S0416 tries to elaborate on the other speaker's turn by giving a concrete example of the phenomenon being discussed.

(4) S0415: there's lots of stars depending on the night
S0475: mm
S0415: depending on whether it's cloudy or not
S0416: ***like*** when we come back from the cinema I always see stars (SX2V/ BNC2014 children).

Use of exemplifier *like* in initial position signals elaboration on previous comments by providing an example (Miller, 2009: 336). I did not observe this kind of elaboration in the Ar-Children corpus, where exemplifier *like* usually precedes a list of members of some specific

category (e.g. of food, school subjects, family members, shopping centres, restaurants or coffee shops) and appears mostly in turn-medial position, as in (5).

(5) 121402: okay what 's your- your favourite food like, outside, ***like*** Mac or Hervy like this,

121401: yeah McDonald's (1214/ Ar-Children).

When they use *like* as an explanator, speakers try to reformulate a previous utterance, or to explain it by supplying an alternative term or repeating a previous utterance (see section 5.2.1). Speakers tend to reformulate when they feel that a previous explanation was insufficient (Müller, 2005: 216). In (6), *work* reformulates *do* to specify that the intended reference is to a job/profession.

(6) 101802: what do you think, your sister will do, when she grow up***, like***, what will she work/ (1018/Ar-Children).

Explanator *like* may alternatively precede repetition of phrases or clauses within a single turn. In this function, the speaker uses almost the same words in their explanation. For instance, the speaker in (7) expresses her feelings about a family trip; by the end of her turn, she repeats the adjective *disappointed*, which she used towards the beginning of her turn.

(7) 102301: and, we just came back, () *like*, uh *like* dis- disappointed, because, *like* uh the, the- the first of the day. it was *like*, so good/ and it ended up bad, so ***like***, I was disappointed/ (1023/ Ar-Children).

Repeated material after elaborator *like* may not always precisely replicate the prior expression of the content; for example, it may have a different syntactic order, as in (8).

(8) S0655: then we managed to fly erm down the staircase so we'd run out of time and ***like*** time is what you run out of (SX4N/ BNC2014 children).

Explanator *like* can alternatively flag repetition at the level of a single word or phrase, perhaps with modification (e.g. adding or changing an adjective in a noun phrase), as in (9). This function is treated in the literature as part of *like* as a lexical focuser. However, in my framework, I identify this function as a subtype of explanator *like* flagging explanation through repetition. This is because *like* with repetition is not used to introduce new information, which would constitute an initiator in my framework (see section 5.2.3.1) or a focuser in previous literature, but to repeat information that has already been established.

(9) 121401: uh what do you like/ in the food,
121402: mm () I like/
121401: that's ***like*** your favourite food,
121402: I like pasta. (1214/ Ar-Children).

I coded 88 instances (0.8 per 100 clauses) of *like* as this sub-type of explanator in the Ar-Children corpus, compared with 132 (0.6 per 100 clauses) in the BNC2014 children subcorpus. In other words, both groups of speakers employ explanator *like* to flag repetition of a single word or phrase, with comparable regularity.

The third type of explanator *like* precedes an attempt to explain a term by giving an alternative. The speaker in example (10) uses *like* to introduce *argument* as an alternative term for *case*.

(10) S0655: >> a per- a lawyer is a person who erm decide erm who can win a case ***like*** an argument (SNCR/ BNC2104 children).

We see, then, that three sub-types of explanator *like* are each present in both corpora. I also observed an additional type of explanator, which I label as *illustrator*. Illustrator *like* is used when the speaker wants to provide more detail on an answer and steps out from a conversational sequence of question-and-answer (Fuller, 2003b: 373). Unlike using *like* as explanator, the speaker is not trying to explain a statement but rather a question. In (11), for instance, the speaker uses *like* to explain her question by giving specific details about the type of answer she is expecting.

(11) 124501: okay. what is the favourite subject. ***like*** you think , that it's easy, and mm very like easy, (1245/ Ar-Children).

Code-switching after elaborator *like* by L2 speakers was observed in the Ar-Children corpus. One explanation for code-switching is that the speaker has a richer vocabulary in their L1 than their L2, such that using their L1 makes them more confident in delivering the right intended message to speakers with the same L1. However, in (12), code-switching occurs as the participants are trying to explain a misunderstanding. Speaker 082402 uses an Arabic expression أنتي شتحبين (*'antī shiṯḥbīn* 'what do you love?') to explain the question after her friend fails to grasp that she is trying to move away from the specific topic of animals.

(12) 082402: what's your favourite pet/

082401: what is pet/ [VOC-laugh]

[…] 082401: oh. I love actually cat/ of course,

082402: ***like*** [VOC-laugh] أنتي شتحبين/

082401: of animals/

082402: yes, what do you love more than animals.

082401: cat.

082402: and me.

082401: oh, and you, let me think about it/ Allah. of course it, Allah / (0824/ Ar-Children).

This use of *like* prior to a code switch that provides explanation occurs three times in the Ar-Children corpus. As such, it is not frequent enough to formalise as an additional subfunction of *like* in my framework. However, it shows that the strategy of "switching languages to explain something" is compatible with being flagged by a DM in the language that the speaker is switching away from. It also adds to our understanding how speakers can utilise code-switching to maintain continuity of discourse.

5.2.3.3. Filler *like*

Filler *like* can mark a self-repair or a search for an appropriate word or phrase. *Like* as a filler is used twice as often in the Ar-Children corpus (3.7 per 100 clause) as in the BNC child subcorpus (1.5 per 100 clause). This result might simply reflect different levels of English fluency leading to planning difficulties for the L2 speakers, and consequently to more self-repair. *Like* as a filler can also function as a politeness indicator, helping speakers avoid sounding too assertive (Andersen, 2001: 229). Example (13) shows *like* as a marker of self-repair in which the speaker repair with a different syntactic form. In (14), by contrast, the speaker flags a false start with *like* before repeating the same syntactic form.

(13) 083302: okay, so / uh I went***, like***, we were *like* in a car / I was at the park, okay, I went there/ I kept hearing noises (Ar-Children/0833).

(14) S0416: >>he did ***like*** he did PE (SX2V/ BNC2014 children).

Like filling a pause during a speaker's search for a word or expression is exemplified by (15); overall, this is the second most prevalent of the DM functions of *like* in both corpora.

(15) 120802: and mm my mum *like* , have, ***like*** mm see ***like***, mm spicy nuggets/ I mm just, mm eat, a little spicy nuggets , (1208/ Ar-Children).

It is notable that 40 out of the 271 of the instances of *like* in the Ar-Children corpus that I coded as a search for an expression are not accompanied by any type of pause. In these 40 examples, *like* occurs clause-medially and carries a raising intonation contour, as in (16). The use of *like* with a rising tone shows that speakers are still in the process of thinking.

(16) I do n't know, do you wanna us go to VR games/ () or ***like/*** somewhere fun (1101/ Ar-Children).

5.2.4. Unclassified instances

Some examples of DM *like* were coded as unclassified, because although they did have a discernible function, that function was not frequent enough to warrant a new category: 23 examples (0.2 per 100 clauses) in the Ar-Children corpus and 36 (0.2 per 100 clauses) in the BNC2014 children subcorpus. One of these functions is use of *like* in turn-initial position prior to questions – accounting for eight examples in the Ar-Children corpus and four in the BNC2014 children subcorpus. In these examples, a speaker appears to be trying to request further elaboration (Miller & Weinert, 1995: 388). For instance, the speaker 080202 in (17) wishes to get additional detail about a typical dish in 080201's country.

(17) 080201: uh in [ANON-place] / I, they have a main dish. but they do n't often eat it there/ () cause often we eat from restaurants, () or *like* other food.
080202: ***like*** what's the main dish about / (0802/ Ar-Children).

One analysis of (17) is that *like* here is an explanator (see section 5.2.3.2). However, as noted in 5.2.1, Miller & Weinert (1995: 390) suggest another analysis. The other possible explanation is that *like* can mark a request for clarification (interpersonal). In a more similar

example (18), speaker 091701 asks her friend a question preceded by *like* as a request for more information regarding a fictional character under discussion.

(18) 091702: I know you love her/ because your mask is pink too.

091701: yeah,

091701: ***like*** explain what is her hair colour, or does she have a boy/ or uh- or children/ (0917/ Ar-Children)

Also coded as unclassified was DM *like* in turn-final position (twice as frequent in the BNC2014 children subcorpus). Diskin (2017) notes that *like* in turn-final position has limited functions (see section 5.2.1) as a retrospective explanator or exemplifier, or in some cases to mark a preceding request for further information (Miller, 2009: 334). For instance, turn-final *like* in (19) could equally be interpreted as a retrospective exemplifier (textual), or as marking a request for clarification (interpersonal). Moreover, Schourup (1985: 36) states that clause-final *like* with falling tone, following questions, reflects a speaker trying to show awareness of the potential inappropriateness of their question. In such a case, *like* contributes to softening the force of the utterance (Miller & Weinert, 1995: 389).

(19) 091801: I went to [ANON-place] in the holidays.

091802: and, where have you gone, ***like***,

091801: in [ANON-place],

091802: yeah, (0918/ Ar-Children).

Speakers can also employ *like* in turn-final position to soften a negative evaluation in the preceding turn (Corrigan, 2015:50), as exemplified in (20).

(20) 113502: and I like also the coffee for Starbucks, and Dunkin /

113501: uh I do n't like Dunkin ***like***,

113502: it's so nice (1135/ Ar-Children).

As noted, *like* is normally a textual DM. Therefore, these rare examples of turn-peripheral *like* with (additionally or alternative) interpersonal function warrant further qualitative examination in the future.

5.3. *So*

5.3.1. *So* in previous literature

Previous studies of *so* have identified functions according to its semantic meanings (Schiffrin, 1987, Redeker, 1990,1991), its pragmatic functions (e.g. Van Dijk, 1979; Fraser, 1990, 1996, 1999) its users, i.e. L1 and L2 speakers (e.g. Buysse, 2007, 2009, 2012, 2014; Müller, 2005; Lam, 2007; Algouzi, 2014, 2021) or following rules of conversation analysis (e.g. Raymond, 2004; Bolden, 2008, 2009). According to Schiffrin (1987), *so* is mainly a textual marker that carries three main functions: main unit marker, causal relation marker and transitional marker. She also identifies interpersonal functions of *so*, including turn-taking marker, request for confirmation, or further explanation.

In Schiffrin's view, *so* as a main unit marker signals a shift back to main or intermediate levels of narrative structure after further details regarding some point have been given. With this function, *so* can mark a conclusive relation between non-adjacent utterances within one or more turns by a given speaker. This type of conclusive function often appears after a *because* clause. Schiffrin (1987: 198) explains that a *because* clause carries subordinate information to a proposition; a subsequent *so* clause serves to reiterate that main proposition, or as she calls it "the main idea unit". This pattern can be schematised as "A *because* B *so* A", and marks a shared relation between speakers (Passot, 2007: 132). Although this pattern involves a reiteration of main idea by the same speaker marking textual function, *so* can be an interpersonal marker that signal speakers' efforts to involve their listeners in conversations. According to Passot's (2007: 132) explanation, this involvement

often happens when a speaker states content in "A" as consensual between speakers. Then, the same speaker offers a "*because* B" utterance in which both speakers reach the proposed consensus. In this case, the "*so* A" utterance works as a reiteration to the main idea and as a shared-background information between speakers. Passot (2007:132) elaborates that speakers use the "*so* A" utterance to flag their supplementary comments on the ongoing conversations which subsequently can guide other speakers to the end point of their argument and encourage them to comment on what has been said. In my framework, I consider this pattern of *so* clause as interpersonal, following Passot (2007).

Schiffrin (1987:202) also differentiates three types of causal relations that *so* can mark: fact-based, knowledge-based and action-based. These three types of causal relations mirror her first three planes of talk: ideational state, information state and action structure, respectively (see section 2.3.2.1). Fact-based relations are between two events or ideas where one is a result of the other on a local or a global level of discourse. Knowledge-based relations, on the other hand, link inferences over ideas. In defining action-based relations, Schiffrin (1987:207) discusses use of DMs *so* and *because* in three action pairs: request and account, compliance and justification, and claim and grounds. Of these, only the request and account action pair involves *so*. A speaker is motivated to use *so* by the other speaker having shared some information (Schiffrin, 1987: 207). This function of *so* entails interaction between interlocutors, because it accompanies a request (Müller, 2005: 82).

The first two types of Schiffrin's (1987:202) taxonomy for causal relations are discussed in earlier work by Van Dijk (1979:453), who offers two interpretation of *so*. First, *so* marks a result of a previously uttered reason or cause, as in "I was sick, ***so*** I stayed in bed" (Van Dijk, 1979:453). This resembles *so* marking a fact-based relation according to Schiffrin (1987). Second, *so* marks a conclusion as in "John is sick. ***so***, let's start" (Van Dijk, 1979:453) – which is what Schiffrin explains as a knowledge-based relation. This latter

function of *so* often appears in turn-initial position to mark a speaker's attempt to infer a conclusion from the content of other speakers' turns (Van Dijk, 1979: 454). As such, *so* with this type of conclusion function is an interpersonal marker of information shared between interlocutors (Buysse, 2012: 1768). *So* also works as an interpersonal marker when it precedes speech acts such as requests or questions (as identified in the third type of Schiffrin's casual relation). Bolden (2008: 308) states that speakers use *so*-prefaced questions to display their knowledge and interest in other speakers' lives. In the present study, I code *so* with requests or questions as a speech act marker to mark requests, clarification or confirmation of previously discussed information.

Fraser (1990, 394; see also Fraser, 1996: 340, 1999: 948), treats *so* as an inferential marker that carry various pragmatic functions. In Fraser's account, *so* can mark a conclusion derived from prior discourse (similar to Schiffrin's main idea marker). Fraser (1990: 395) also argues that *so* can carry pragmatic functions beyond its meaning as a marker of result. However, Fraser (1990) does not consider in detail how contextual factors may affect the pragmatic interpretation of the examples he uses in his account (Müller, 2005: 64). For instance, Fraser proposes that in the example "Teenage son: The Celtics have an important game today. Disinterested parent: *So*?", *so* functions to challenge the hearer to provide the reason why they uttered what they just said. Müller (2005: 64), argues that *so* can carry another interpretation beside that of Fraser's. In uttering *so*, the parent shows that they are questioning the implication of what their son has said (Müller, 2005: 64). This interpretation depends mostly on background information, for instance, about the team and the relationship between the boy and the team. My interpretation for Fraser's (1990: 395) example is that *so* works as a turn-management marker, since the speaker uses *so* to relinquish the conversational floor and gives the other speaker a chance to persuade their discussion. The parent in this example may acknowledge that their son is about to discuss the matter of the

game. Hence, in using *so*, the parent is returning the responsibility to their son to complete the matter under discussion (Raymond, 2004:209). Interpreting such an example of standalone *so* is not straightforward, as in most cases it requires further contextual information (Müller, 2005).

Buysse (2012) observes the various functions of *so* under three domains (ideational, interpersonal and textual) in a corpus of Dutch speakers of English as a L2, in comparison to L1 English speakers. *So* works as a marker of result in the ideational domain and as a marker of conclusion in the interpersonal domain (as discussed in Van Dijk, 1979). With regard to the interpersonal domain, Buysse finds that turn-final *so* can signal that the speaker wishes to claim the floor. In the textual domain, Buysse (2012) illustrates that *so* functions as a marker to introduce a new section of discourse, a new sequence or an elaboration. According to Buysse (2009), *so* functions as an elaborative marker used when speakers append specific details to the information they have given. *So* can also indicate a shift to a higher unit of discourse; this is Buysse's version of Schiffrin's (1987) main unit marker, or "boundary marker", per Müller (2005: 80). Finally, Buysse (2012) notes that *so* can operate as a filler, in which case it can signal self-repair.

Buysse (2012) concludes that L2 speakers overuse some of the textual functions of *so* (i.e. elaboration and filler) in comparison to L1 English speakers. Similarly, Müller (2005) finds that German speakers learning English as L2 overuse *so* with textual functions, in comparison to L1 American speakers. Like Müller (2005), Algouzi (2014, 2021) reports that adult speakers (Saudi Arabian undergraduates) learning English, as represented in the Arabic component of the Louvain International Database of Spoken English Interlanguage corpus (Ar-LINDSEI), employ *so* mainly as a textual marker, but that they underuse *so* in comparison to L1 adult speakers.

Other studies report that L2 speakers tend to use *so* with a similar or higher frequency than L1 speakers. For instance, Lam (2007) observes the use of DM *so* in the Hong Kong Corpus of Spoken English (HKCSE) in comparison to a customized subcorpus from the BNC1994, representing L2 and L1 English speakers respectively. Lam finds the use of *so* to be similar in both datasets. Following Lam (2007), Weirong (2017) investigates the use of DM *so* in the Asian corpus of English (Kirkpatrick, 2010) and the BNC, finding it to be more frequent in the speech of English learners in the ACE than in the L1 speech in the BNC.

In sum, previous studies of *so* as a DM, which focus on adult L1 and L2 speech, have shown that its use goes beyond its core semantics of result to multiple functions across different functional domains. Moreover, the frequency of *so* in L2 speech differs from its frequency in L1 speech.

5.3.2. Frequencies of the functions of DM *so*

There are 642 and 952 occurrences of *so* in the Ar-Children corpus and the BNC2014 children subcorpus respectively. Table 5.3 shows its frequencies as a DM and non-DM.

Table 5.3. *Raw & relative frequencies (per 100 clauses) of* so *in the Ar-Children corpus and the BNC2014 children subcorpus*

Corpus	Total use of *so*	DM use	Non-DM use
Ar-Children	642 (6.0)	322 (3.0)	320 (3.0)
BNC2014 children	952 (4.3)	622 (2.8)	330 (1.5)

DM use accounts for exactly half the examples of *so* in the Ar-Children corpus and approximately two-thirds in the BNC2014 children subcorpus. The relative frequencies for *so*

as a DM are similar in the two corpora. Table 5.4 shows the frequencies of *so* in both corpora according to the framework of DM functions (see section 3.5.2).

Table 5.4. *Raw and relative (per 100 clauses) frequencies and percentages of* so *with different functions in the two corpora*

Function		**Ar-Children**			**BNC2014 children**		
		N	Per 100	%	N	Per 100	%
Textual	initiator	64	0.6	19.9	21	0.1	3.4
	elaborator	43	0.4	13.3	120	0.5	19.3
	filler	40	0.4	12.4	78	0.3	12.5
	transitional	110	1.0	34.2	168	0.7	27.1
	Total	**257**	**2.4**	**79.9**	**387**	**1.7**	**62.3**
Interpersonal	speech act	22	0.1	6.8	83	0.3	13.3
	conclusion	25	0.2	7.8	79	0.3	12.7
	turn management	18	0.2	5.6	60	0.3	9.6
	Total	**65**	**0.6**	**20.2**	**224**	**1.0**	**35.6**
Unclear		0	0.0	0	13	0.1	2.1
Total		**322**	**3.0**	**100**	**622**	**2.8**	**100**

So is predominantly a textual marker in both corpora; this accounts for four-fifths of instances in the Ar-Children corpus (79.8%) and about two thirds in the BNC2014 children subcorpus (62.3%). In the Ar-Children corpus, the two predominant textual functions are the transitional and initiator functions. By contrast, speakers in the BNC2014 children subcorpus employ *so* mainly with transitional and elaborator functions. The proportions of *so* as filler

are similar in the two corpora. In the interpersonal domain, Arabic-speaking children employ *so* to serve conclusion and turn management functions twice as often as speech act function. The relative frequencies of *so* with interpersonal function in the BNC2014 children subcorpus are similar for all subfunctions (0.3 per 100 clauses).

The differences between the frequencies of the functions of *so* in the two corpora were tested for significance using the Log-Likelihood test. To perform the test, I followed the same steps used in testing the data on functions of *like* (see section 5.2.2). The functions of *so* differ significantly between the two corpora at the $p<0.001$ level ($p = 2.2x10^{-16}$, LL= 88.995, df = 7).

The next section presents the textual and interpersonal functions of *so* with examples from the two corpora.

5.3.3. Qualitative analysis of textual functions of *so* in context

5.3.3.1. Initiator *so*

So as an initiator is six times more frequent in the Ar-Children corpus (0.6 per 100 clauses) than in the BNC2014 children subcorpus. This function signals the beginning of a new topic within an ongoing discourse after a pause or after a request from other speaker(s) to participate with a new topic. In addition, I also observed that speakers in the Ar-Children corpus employ the initiator *so* to mark a new topic at the beginning of their conversation as in (21). I could not code such instances of *so* in the BNC2014 children subcorpus. This is because unlike the Ar-Children corpus, the transcripts in the BNC2014 children subcorpus are not synced to the beginning of a specific conversation/task, making it hard to recognise the start of conversation as opposed to start of new topic. Thus, I could not detect this subfunction of initiator *so* precisely in the L1 data.

(21) 102301: hey,

102302: hi,

102301: hi. ***so*** let's let's talk about the holidays/

102302: okay, (1023/ Ar-Children).

This use of initiator *so* is sometimes accompanied with DM *okay*, as in (22).

(22) 120801: *okay*. ***so***, we should start with family. how many siblings/ do you have.

(1208/ Ar-Children)

The speaker in this example uses *okay* to show that she is willing to take the lead in the conversation and start the discussion (see section 6.3.4.1). Then, *so* flags the initiation of her suggestion for the first topic of discussion. Müller (2005) observes the use of initiator *so* (boundary marker in her analysis) when accompanied with *okay* at the beginning of conversations. According to Muller's (2005) argument, speakers tend to use *okay* to express their agreement with the researcher who has explained the conversational task (see section 6.2.4). Then, they use *so* to introduce their first topic of discussion to the other speaker. In my data, I observe that Arabic-speaking children employ this use of *so* to initiate the first topic of discussion after *okay* in four examples, only. I also observed that participants use the construction *okay+so* in 16 instances amid their conversations. In these examples, speakers use *okay* for other functions such as filler or turn management functions (see section 6.2). However, they use *so* after *okay* to signal their willingness to initiate a new topic as illustrated in (23).

(23) 084602: my mum go out/ and then she didn't see it/ him- he- she did n't see him / then the next time, uh she found him with the security/

084601: security/

084602: security, and then we go to the hotel, so. this is my scary story. my brother/

084601: ***okay, so*** this is a story about my sister,

084602: okay (0846/ Ar-Children).

5.3.3.2. Elaborator *so*

Elaborator *so* appears slightly more frequently in the BNC2014 children subcorpus (0.5 per 100 clauses) than in the Ar-Children corpus (0.4 per 100 clauses). In example (24), speaker 101801 uses *so* to flag an upcoming explanation of initiating her previous utterance "my mum has Airpods".

(24) 101801: my mum has AirPods/ ***so*** I always take them from her. and I just use them, like I ask her, can I use them, and then I use them for like hours, I call my friends with them, I listen to music, watch YouTube, (1018/ Ar-Children)

Speakers also employ elaborator *so* prior to exemplification. Buysse (2009: 84) observes that speakers often utilize elaborator *so* alongside other exemplification markers such as *for example* or *like*. However, this is not the case in the Ar-Children corpus, where speakers use elaborator *so* alone to signal exemplification. For instance, in example (25), speaker 125701 employs elaborator *so* to flag an example of what she has learned in the gymnastic class.

(25) 125701: (…) uh I go to like/ a gym gymnastics/ and, I learn some stuff/ ***so,*** I learn the handstands. (1257/ Ar-Children)

In the BNC2014 children subcorpus, speakers use elaborator *so* to explain previous information (n= 111) and offer further exemplification (n=11). In example (26), speaker S0416 uses *so* to flag their turn as an explanation of their previous one, and in so doing signals continuity in the conversation.

(26) S0416: and you you are a vault hunter

S0417: mm

S0416: okay ? ***so*** Handsome Jack tries to kill you because obviously you 're his contester you 're the one you 're the one (SDFR/ BNC2014 children)

5.3.3.3. Filler *so*

In the Ar-Children corpus, speakers use *so* as a filler to repair their speech in 4 instances (0.0 per 100 clauses) and as a planning function in 36 instances (0.3 per 100 clauses). The filler *so*, in the BNC2014 children subcorpus, marks a self-repair function in 15 instances (0.1 per 100 clauses) and a planning function in 66 instances (0.3 per 100 clauses). Example (27) shows the use of *so* as a self-repair marker to flag the speaker's reformulation of her previous utterance:

(27) 082101: […] mm I do n't really remember the two. but they were like going to a place/ and then, uh they wanted- they were at the- at their home. ***so***. when they were at their home . (0821/ Ar-Children)

In the Ar-Children corpus, I coded *so* as a filler flagging a planning function when it is accompanied with filled or unfilled pauses, pronounced with a rising tone and prolonged vowel, or when it appears as a standalone marker in a single turn. In example (28), the speaker uses *so* as a standalone marker to signal that she needs time to think and does intend to say something.

(28) 084601: gold. my colour is rose gold, but like if it's colour, of like clothes or something, it s going be uh.violet or pink or baby blue.

084602: ***so***/ ()

084601: you're thinking. (0846/ Ar-Children).

In some cases from the Ar-Children corpus, *so* occurred with a falling tone but accompanied with filled pause signals a planning function, as in (29).

(29) 110502: uh okay, I 'm going to tell you a story. mm () uh ***so.*** the story is about my- the story is about mm my pet mm my pet- I have a pet named [ANON-name], (1105/ Ar-Children)

5.3.3.4. Transitional *so*

As a transitional marker, *so* conveys either sequential function, topic resuming/changing function, or resultative function. Redeker (1990: 373) argues that *so* marks sequential relations in the chain of a narrative parallel to the use of *then*. Speakers in the Ar-Children corpus employ *so* in this way eighteen times, as in (30).

(30) 090801: there is someone who is sitting with his friend in the house / he hears the doors open and closed. ***so***, they go and see no one / . (0908/ Ar-Children)

So as a sequential marker is sometimes prefaced with *and* to signal topic continuity (Koops & Lohmann, 2022: 163). This use of *and so* is found only in the BNC2014 children subcorpus. Out of the 36 examples of sequential *so*, six are prefaced with *and*, as in (31).

(31) S0416: er it 's like erm a robot that has like weapons on it well or you can have a gigantic baby I du n no y- you can build it we ma- we we made erm a car erm but the other person drives it and you have to get in () and ***so*** I got in and then he just drove me off a cliff (SR87/ BNC2014 children).

In the Ar-Children corpus, transitional *so* indicates topic resuming function in 49 instances (1.0 per 100 clauses). This use of transitional *so* flags a return to the main point in the speaker's narrative. For instance, the main line of the narrative around example (32) is about the speaker's sister being allergic to cats, despite being keen to have one. But at the start of (32), the speaker has shifted to expressing her and her sister's feelings on pets. The use of *so* marks a return to the narrative to complete the story.

(32) 693002: I 'm scared of cats. but then , when we got a cat , like , for my sister / because , she 's a cat person / and I 'm like , a dog puppy person , ***so*** , she begged my mom and dad to buy , a cat , and then , they buyed it for her, (6930/ Ar-Children)

As a transitional marker, *so* can also signal a resultative relation between two discourse units, as it "brings a sequence or a turn to close" (Buysse, 2012: 1773). *So* as a resultative marker could in many cases be paraphrased by *therefore* (Buysse, 2012:1765). The resultative function of transitional *so* appears in 43 examples in the Ar-Children corpus (0.4 per 100 clauses) and 91 examples in the BNC2014 children subcorpus (0.7 per 100 clauses); one example is given in (33).

(33) 085301: this summer, I'm going to Miami. I'm going to Miami. mm , we ca n't stay there for very long, because my cousin's getting married, […] we might go before , or after, we're still not very sure,

085302: I feel like. it would be better to go before, ***so***, you could get some dresses, (0853, Ar-Children).

5.3.4. Qualitative analysis of interpersonal functions of *so* in context

5.3.4.1. *So* as a speech act marker

As a speech act marker, *so* appears in 22 instances in the Ar-Children corpus (0.1 per 100 clauses) and in 83 instances in the BNC2014 children subcorpus (0.3 per 100 clauses). *So* functions as a speech act marker to signal a request for clarification or confirmation about another speaker's statement. This use of *so* often appears with questions. As a seeking for clarification marker, speakers employ *so* to request more information about the topic. In example (34), speaker 084602's utterance "we did not find him" motivates the other speaker, 084601, to request further information about when or how they found her brother. *So*, in this

example, flags her request for more information about 084602's story as a way of showing her engagement.

(34) 084602: and then/ we didn't find him.

084601: ***so***. how did you find him.

084602: my mum go out/ and then, she didn't see it/ him- he- she did n't see him/ then the next time, uh she found him with the security. (0846/ Ar-Children)

As a marker of request for confirmation, *so* flags a speaker's attempt to confirm their understanding of given information. In example (35), speaker 120801 requests confirmation of her attempt to help her friend articulate the name of a brand of chips. This use of *so* expresses uncertainty on the part of the speaker, which is underlined by the use of a rising tone at the end of the utterance.

(35) 120802: my favourite chips is mm like, this thing, that have been in ads. like, I dunno what's called.

120801: you just shaped a triangle, with your hand,

120802: like this/

120801: ***so*** you like Doritos/

120802: yeah/ (1208/ Ar-Children)

As a speech act marker, *so* may also preface a request from a speaker to change turns as in (36). This subfunction is infrequent in the Ar-Children corpus with only four examples of *so* flagging a request for turn change.

(36) 124502: yeah. okay , ***so*** can you now, ask me the question/ (1245/ Ar-Children)

I observed that in the BNC2014 children subcorpus, speakers also utilize *so* as a request marker to request permission to perform a physical action, as in (37).

(37) S0655: I feel like I 'm worry that if I have another main I 'm going to blow up ***so*** can I please get down? (SP45/ BNC2014 children).

5.3.4.2. *So* as a marker of expressing conclusion

As a conclusion marker, *so* is twice as frequent in the BNC2014 children subcorpus (0.6 per 100 clauses) as in the Ar-Children corpus (0.3 per 100 clauses). I observe two subfunctions of *so* as conclusion marker in both corpora: a conclusion marker to background information and as a marker for reiteration. Although these two functions may appear to be textual functions, I opt to code them as interpersonal following previous research (see section 5.2).

The first subfunction of *so* as a conclusion marker is to mark a conclusion after speakers have had all the background information from the other speaker(s). For instance, in example (38), speaker 080202 comes to the conclusion that the appropriate label for the games her friend prefers is "VS games". Speaker 080201 shows a hedged agreement "something like that" to her friend's labelling to emphasise that these are 080202's words and not her own.

(38) 080201: mm I like, like/ a genre of shooting and violence.

080202: ***so*** VS games. okay,

080201: something like that you can say. (0802/ Ar-Children)

I also observe that speakers in the Ar-Children corpus often employ a level or falling tone unit to the following utterance in using *so* as a conclusion marker, similar to the previous example. In the BNC2014 children corpus, the conclusion *so* marker is often followed by utterances that do not receive any confirmation or disagreement from other speakers. For instance, example (39) shows how speaker S0655 comes to the conclusion that the character name is "Lisa", after considering the different propositions they have already stated. The

other speaker did not agree or disagree using any dis/agreement words (e.g. *yeah, yes, no*) with S6055 but rather produced another proposition that semantically opposes their statement.

(39) S0654: her () it's somebody who can look after her anyway

S0655: erm she is a vampire and she was painted and her picture was the Mona Lisa () ***so*** maybe her name is Lisa

S0654: --UNCLEARWORD name is Lucy

S0653: >>maybe (S839/ BNC2014 children).

So can also mark a conclusion with a reiteration of the main idea (as discussed in section 5.2). The reiteration of main idea can occur between utterances, within a single turn or sometimes over a series of turns, uttered by the same speaker as a three-stage structure (A *because* B *so* A). In the Ar-Children corpus, *so* marks a conclusive function by the same speaker in single turns. For instance, in example (40), the speaker reiterates her first proposition after providing some supplementary information to elaborate on her narrative. The reiteration guides the other speaker to the main point of narrative that is presented as a conclusion with *so*.

(40) 083401: they, uh, they show me my brother. I was hugging him, because I miss him so much. () ***so*** I was hugging him. (0834/ Ar-Children)

The three-stage (A *because* B *so* A) conclusion function of *so* is also employed by speakers in the BNC2014 children subcorpus. However, unlike the examples in the Ar-Children corpus, it is sometime used over a series of turns, as in example (41). Speaker S0622 uses *so* to flag their reiteration of the main idea (i.e. not going on the caravan) after giving the reasons for that.

(41) S0622: and --ANONnameM for the last couple of miles of him saying this is too far this is too far cos it was taking us er () a good while while or three quarters of an hour to get there

S0525: oh right

S0622: and er s- --ANONnameM says I well well I 'm not gonna stay in the caravan if it's gonna be this () this far

S0525: >>this far away?

S0622: >>because we () we'd have to go back early in the morning for half seven to c- to start

S0525: >>yeah

S0622: the wagons rolling and that sort of thing

S0525: >>yeah yeah

S0622: >> ***so*** we're not gonna well we won't go to this and --ANONnameM was going (SA2J/ BNC2014 children).

5.3.4.3. *So* as a turn management marker

In this function, *so* flags an attempt to retake the conversational floor so that the speaker can continue or elaborate on their previous turn. This function differs form the initiator functions (section 5.3.3.1) in that speakers using this function appear to take responsibility for leading the conversation, following a unsuccessful attempt from another speaker to participate in the ongoing interaction. In the initiator function, *so* flags a new topic that is initiated by the speaker following the other speaker's request to take over the conversational floor or following the researcher's request to start the conversation. Thus, in using *so* as an initiator, speakers participate in the continuity of discourse rather than trying to take over the floor from the other speakers.

In example (42), speaker 120801 employs *so* to flag her attempt to hold the floor after her friend's unsuccessful attempt to contribute.

(42) 120801: every time I talk , with my friends/ I go to a new topic from one topic, it's so hard finding the first topic,

120802: yeah, [VOC-laugh] mm ()

120801: ***so*** we find the first topic, and they will lead to other topics [VOC-laugh] (1208/ Ar-Children)

Speakers sometimes also employ *so* as a turn management marker to flag an attempt to relinquish the conversational floor to another. This happens when *so* appears in turn-final position, and sometimes followed by a discernible pause. For instance, in example (43), speaker 084801 ends her turn with *so* after completing a story. Speaker 084802 appears to recognize this as an indication that speaker 084801 has finished her turn, and takes over the floor. This example of *so* having a rising tone may also encourage the other speaker to take the conversational floor.

(43) 084801: you know / if- if I never changed seats, I would n't have met you. if-if it just , if I just stayed/ in the same seat. same table / I would've been friends, with other people. If I did n't change seats, I would n't have been friends with you, ***so***/

084802: yeah. I 'm really happy that we're friends. (0848/ Ar-Children)

The interpersonal functions of *so*, including turn management, are more frequent in the BNC2014 children subcorpus than in the Ar-Children corpus. This suggests that L1 speakers are more aware of using this DM to manage their relationship with other speakers during their interactions than L2 speakers.

5.5. Summary

This chapter presented a functional qualitative analysis of *like* and *so* in my two corpora. This qualitative analysis of these two DMs adds to the response to the third RQ of the present study (see section 1.3). DMs *like* and *so* are among the most frequent in my data and in both corpora, they predominantly perform textual functions.

In section 5.2.1, I presented an overview of work on the functions of *like* in the previous literature. Then, I discussed the frequencies of these functions in the two corpora. I found textual functions of *like* to be three times more frequent in the Ar-Children corpus than in the BNC2014 children subcorpus. In section 5.2.3, I presented an in-depth qualitative analysis of the various textual and interpersonal functions of DM *like* in my corpora. As an elaborator, *like* fulfils different functions, including acting as an exemplifier, an explanator or a focuser. As a filler, *like* flags the following utterance as being a self- repair or search for appropriate expression marker. This is in accordance with findings from earlier studies (e.g. Andersen, 2001; Müller, 2005). Uses of *like* preceding code-switched utterances or as a search for appropriate expression with a rising intonation, as observed in the Ar-Children corpus, appear to extend the documented functions and warrant further investigation. In section 5.2.4, I argued that *like* can carry an interpersonal function in initial and final positions. However, due to the rarity of examples of *like* in peripheral positions, I could not generalise from this observation.

In section 5.3, I presented the functional analysis of *so* in the two corpora. In 5.3.1, I once again reviewed relevant prior literature related to *so* as a DM. In 5.3.2, a quantitative analysis revealed that *so* is predominantly used with textual functions in both corpora. This finding aligns with observations made in previous studies (e.g. Lam, 2007; Buysse, 2012). While it functions mainly as a transitional and initiator marker in the Ar-Children corpus, it functions mainly as a transitional and elaborator marker in the BNC2014 children subcorpus.

In the two corpora, children employ initiator *so* to mark new topics. They also employ the elaborator *so* to mark explanation and exemplification. Likewise, in both corpora, filler *so* signals planning for an upcoming utterance. Transitional *so* carries various subfunctions, including topic resuming and resultative functions.

The analysis shows that *so* as an interpersonal marker (see section 5.3.3) occurs more frequently in the BNC2014 children subcorpus than in the Ar-Children corpus. In the interpersonal domain, *so* works as a speech act marker to introduce a request for clarification or confirmation. Additionally, *so* can carry a conclusion function in both corpora. *So* can also mark a turn management function, flagging a speaker's attempt to hold or relinquish the conversational floor. This, again, is observed in both corpora.

This chapter has discussed the functions of DMs *like* and *so* in my data. Both are predominantly textual in function, although do – less frequently – exhibit interpersonal functions. The following chapter continues the discussion of the most frequent DMs in the two corpora by applying the same analysis to three more DMs, namely *okay*, *oh*, and *yeah*.

Chapter 6: *Okay, Oh* and *Yeah*

6.1. Overview

This chapter discusses the functions of three of the most frequent DMs in the Ar-Children corpus and the BNC2014 children subcorpus: *okay, oh* and *yeah*. These three DMs predominantly serve interpersonal functions in both corpora (see section 4.4). The qualitative analysis of these markers provides part of the answer to the third RQ: *what are the functions of the most frequent DMs in the speech of Arabic school-age children? Are there any differences in their use of DMs from native English-speaking children?*

In this chapter, each DM is addressed in subsequent sections: 6.2 for *okay*, 6.3 for *oh* and 6.4 for *yeah*. In each discussion of a DM, the first subsection reviews related literature, drawing upon previous research to establish the basis for understanding that DM's functions in spoken conversations and in SLA. The second subsection presents the frequencies of the DM's functions in both corpora. In the last subsection, I report the qualitative analysis of the functions of the DM in the speech of L1 and L2 children.

6.2. *Okay*

6.2.1. *Okay* in previous literature

Several studies have examined the functions of *okay* in formal settings, such as police interviews (e.g. Gains, 2011) or academic settings such as seminars (e.g. Rendle-Short, 2000) or university lectures (e.g. Schleef, 2005; Othman 2007, 2010; Al Makoshi, 2014). Other studies have employed corpus methods such as contrastive analysis to compare the use of *okay* in different languages, addressing such pairs as English and Spanish (Amador Moreno et al., 2013), English and Irish (O'Keeffe & Adolphs, 2008) and English and Chinese (Yeh &

Huang, 2016). Each of these studies reports that *okay* can carry several functions at the pragmatic and discourse level.

Early studies describe *okay* (alternatively, *ok*) as mainly a textual marker that marks brackets of a unit of discourse (e.g. Sinclair & Coulthard, 1975; Merritt, 1978; Goffman, 1981; Condon, 1986) or to mark a transition to a closure (e.g. Schegloff & Sacks, 1974; Beach 1993). In their observations of teachers' and pupils' talk, Sinclair & Coulthard (1975: 22) find that *okay*, along with other DMs including *well*, *now* and *right*, work as boundary markers that signal the end of one communicative task and the beginning of another. In line with that observation, Merritt (1978) observes two primary functions of *okay* as a transitional marker in response to actions in service encounters (e.g. between servers and costumers). Firstly, *okay* punctuates sequences of communication (Merritt, 1978:15). For instance, when a server utters *okay* and moves to perform an action (e.g. to bring something to a customer), *okay* marks the transition from verbal to non-verbal action. Secondly in Merritt's (1978) account, *okay* marks a closure in a customer's speech. When used by customers, *okay* serves as a signal to release the server from any further obligation. Therefore, *okay* can mark a speaker's intention to show not only their acknowledgment but also their satisfaction, acceptance or disappointment regarding the service provided to them.

Condon (1986) studies the use of *okay* in discussions among families organizing holiday trips. The results of this study are congruent with earlier observations of *okay* as a transition marker between units of discourse. Moreover, *okay* serves as a linking marker in decision sequences as a first pair part in the conversation (Condon, 1986: 95), for example: "*OK*. The second day. How long do we want to stay in London or England?" (Example 12 in Condon 1986: 89).

The studies considered so far focused on the textual functions of *okay*. However, other researchers subsequently observe many interpersonal functions. Using *okay* to interact

with another speaker and convey emotional status demonstrates that the marker can signal interpersonal alignment. Beach (1993: 341) illustrates that speakers employ *okay* as turn-opener and turn-transitional device used to maintain coherence in discourse. But *okay* can mark more than an acknowledgment or transition. A single turn *okay* with rising tone can work as a tag question (Beach 1993, 330). *Okay* can also be a response marker, signalling approval of other speakers' turns, or to express approval of an answer given by an interlocutor, in a structure such as: "A: You want me to bring you anything? B: No. A: *Okay*" (example 7 in Beach, 1993: 331). In a recent study, Beach (2020) observes that analysing *okay* with consideration of its prosodic contour reveals that speakers use *okay* to show their understanding of, agreement with or disagreement with the matter under discussion. In these cases, *okay* has the interpersonal function of indicating the speaker's alignment with prior discourse.

Okay functions differently according to its position. Huddlestone and Fairhurst (2013) observe this to be the case by examining data from the South African spoken component of the International Corpus of English. *Okay* in turn-initial position tends to have a turn-taking function. In turn final position, *okay* works either as a closure marker, to indicate that the speaker has finished their turn, or as a comprehension check marker (similar to a tag question: Beach, 1993: 335). In turn-medial position, *okay* can carry various discourse functions, including pausing to gain time for thinking and transitioning toward indirect speech within a single turn (Huddlestone and Fairhurst, 2013: 102).

Researchers have also considered the functions of DM *okay* in the speech of adult L2 learners compared with that of L1 speakers. House (2013: 64) observes the use of *okay* in the speech of adult L2 German learners of English. The results show that undergraduate German learners of English as a L2 employ *okay* to serve a variety of functions, including a starter, a comprehension check marker, a receipt for approval, a linking device and a pre-close or

closing marker. That is, adult L2 learners are able to use *okay* to perform functions similar to functions found in their L1. L2 researchers are also interested in the frequency of *okay* in the speech of L2 learners. Arya (2022) observes that *okay* is one of the most frequent DMs in the speech of Thai learners of L2 English. In Ayra's data, *okay* marks various functions in the speech of L2 learners, such as a filler and a marker to introduce new topics. The interpersonal functions of *okay* – including turn-taking and turn holding – are the most prevalent functions of *okay* observed in the speech of L2 Thai learners of English (Arya, 2022).

Okay has also been studied in the communications of L1 children, as an indicator of their level of pragmatic competence. Kyratzis & Ervin-Tripp (1999: 1334) find that young English-speaking children (between 4 to 7 years old) employ *okay* as a boundary marker between parts of narratives. Moreover, the children used *okay* along with DM *but* to preface negotiations with other children. Consistent with Kyratzis & Ervin-Tripp's results, Escalera (2002: 30) finds that young English-speaking children (up to five years old) are able to use *okay* to perform social action and mark participation in the conversation (interpersonal functions in my framework). Overall, it seems that children with L1 English are able to use *okay* as a DM from a young age.

In summary, *okay* has been shown to have various functions in different contexts and positions, and these functions extend beyond its use as a textual marker.

6.2.2. Frequencies of the functions of DM *okay*

As previously undertaken for other DMs, all tokens of *okay* were extracted from the Ar-Children corpus and the BNC2014 children subcorpus, and their discourse functions determined. Table 6.1 presents the frequencies of *okay* overall and as DM and non-DM in the two corpora.

Table 6.1. *Raw & relative frequencies (per 100 clauses) of* okay *in the Ar-Children corpus and the BNC2014 children subcorpus*

Corpus	**Total use of *okay***	**DM use**	**Non-DM use**
Ar-Children	560 (5.2)	548 (5.1)	12 (0.1)
BNC2014 children	448 (2.0)	397 (1.8)	51 (0.2)

After excluding non-DM occurrences, *okay* was qualitatively coded for DM functions according to the framework outlined in section 3.5.2. Table 6.2 presents these functions, according to the textual and interpersonal domains.

Table 6.2. *Raw and relative (per 100 clauses) frequencies and percentages of* okay *with different functions in the two corpora*

Function		**Ar-Children**			**BNC2014 children**		
		N	Per 100	%	N	Per 100	%
Textual	initiator	7	0.1	1.3	0.0	0.0	0.0
	filler	95	0.9	17.3	20	0.1	5.0
	transitional	135	1.3	24.6	21	0.1	5.3
	Total	**237**	**2.2**	**43.2**	**41**	**0.2**	**10.3**
Interpersonal	turn manage.	96	0.9	17.6	109	0.5	27.5
	confirmation	97	0.9	17.7	81	0.4	20.4
	speech act	43	0.4	7.8	33	0.1	8.3
	backchannel	75	0.7	13.7	116	0.5	29.2
	Total	**311**	**2.9**	**56.8**	**339**	**1.5**	**85.4**
Unclassified		0.0	0.0	0.0	17	0.1	4.3
Total		**548**	**5.1**	**100**	**397**	**1.8**	**100**

Although *okay* is predominantly an interpersonal marker in both corpora, a lower proportion (56.8%) of instances carry interpersonal functions in the Ar-Children corpus than in the BNC2014 children subcorpus (85.4%). As an interpersonal DM, *okay* is mostly used as a turn management marker or backchannel in both corpora. The most prominent textual function in the Ar-Children corpus is the transitional function.

The differences across the two corpora between the frequencies of the functions of *okay* were tested for significance using the Log-Likelihood test, applying similar steps to those used to test the equivalent frequencies for other DMs (see section 5.2.2). According to

the Log-Likelihood test, the differences between the two corpora for the functions of *okay* were highly significant at the $p<0.001$ level ($p = 2.2x10^{-16}$, LL= 441.62 and df = 8). This result shows that the frequencies of *okay*'s functions differ between the two datasets.

6.2.3. Qualitative analysis of textual functions of *okay* in context

Only three of four textual functions in my framework appear in the analysis of *okay*, namely the initiator, filler and transitional functions.

6.2.3.1. Initiator *okay*

The initiator function of *okay* marks the beginning of a new conversation or the first conversational topic. DM *okay* as initiator is found only in the Ar-Children corpus, in 7 instances. In (1), *okay* marks the first turn in the conversation.

(1) 120801: ***okay.*** so, we should start with family. how many siblings/ do you have.

120802: I have mother/ and father and brothers/ and, mm their also me. and, uh

(1208/Ar-Children)

A possible explanation for the presence of initiator *okay* function in Ar-Children but not BNC2014 children was outlined for DM *so* in (5.3.3.1), namely that it relates to being unable to detect the start of conversations in the BNC2014 children subcorpus (unlike the Ar-Children corpus).

6.2.3.2. Filler *okay*

Use of filler *okay* appears more frequently in the Ar-Children corpus (0.9 per 100 clauses) than in the BNC2014 children subcorpus (0.1 per 100 clauses), although the filler functions are similar in both. For instance, in the Ar-Children corpus, there are 14 instances of *okay* which mark a speaker's planning. The use of *okay* to mark a self-repair is less

frequent in the BNC2014 children subcorpus than in the Ar-Children corpus. I will discuss this further in section 7.3.

Out of 95 instances of *okay* in the Ar-Children corpus, 90 demonstrate the planning subfunction. As a planning filler, *okay* co-occurs with filled and unfilled pauses and sometimes with other DMs such as *so* and *yeah.* The examples of *okay* in (2) are produced during a planning process, as speaker 124502 struggles to find the name of a movie.

(2) 124502: say any movie , that you- that you watched , (..) yeah,

124501: okay. it's uh called , () the- ***okay/***

124502: yeah/

124501: ***okay,*** (..)

124502: the name of this movie is what/

124501: mm خلاص (khlāṣ 'okay') it's the moon, yeah I like it so much. (1245/ Ar-Children)

The use of filler *okay* to mark a self-repair is rare in the Ar-Children corpus (0.1 per 100 clauses). An example of this subfunction appears in (3), in which *okay* flags a resumption after a self-repair.

(3) 123001: brother. yeah, they go to- to the cats. ***okay/*** they go to buy the cats/ (Ar-Children, 1230)

6.2.3.3. Transitional *okay*

The transitional function is the most frequent textual function of DM *okay* in the Ar-Children corpus (1.3 per 100 clauses). In contrast, it is rare in the BNC2014 children subcorpus (0.1 per 100 clauses). This function is discussed in detail in the literature (Merritt,

1978; Condon, 1986; Beach, 1993, 2020; House, 2013 among others). As a transitional marker, *okay* signals a transition between topics or between aspects of the same topic. It also marks a speaker's attempt to announce closure of their contribution – either with respect to their turn or the whole conversation. Transitional *okay* can also flag quoted speech. The quotative function is rare in both corpora (seven examples in Ar-Children corpus and four in BNC2014 children subcorpus).

Transitional *okay* marks a topic change in 105 examples in the Ar-Children corpus (0.9 per 100 clauses), where such topic changes are often guided by the card box of topics (see section 3.2.1). Such a change may also occur between sub-topics within one general topic, with no change of conversational task. For instance, speaker 121402 employs *okay* to move smoothly between aspects of one main topic (food).

(4) 121401: do you like ice cream chocolate or,

121402: vanilla/

121401: yeah,

121402: I-I like vanilla. ***okay/*** what 's your favourite chips, (1214/ Ar-Children)

Transitional *okay* can also mark the closure of a series of thoughts (or the entire conversation), as illustrated in (5).

(5) 111602: and then we went to the movie, and we watched / mm like , mm yeah Barbie and , the-and the- and some animation, cartoons, like this , and , uh ***okay.*** the third day / I also go to , the pool/ with my cousin,

111602: wow / (1116/ Ar-Children)

As a marker of closure, *okay* can signal in advance the ending of the ongoing conversation – what Schegloff & Sacks (1974: 80) identify as a pre-closing function. In the

Ar-Children corpus, speakers use *okay* to signal that the end of the conversation is nearly coming, as in (6).

(6) 091702: ***okay***/ guess it's time to go, bye/

091701: bye, (0917/ Ar-Children)

Transitional *okay* can also highlight a transition to quoted speech, as in (7).

(7) 101202: […] that uh her stepsisters / say that your father is in- uh your father is sick. so, uh Belle was- uh was very nervous about-about uh her dad, so, she told the beast, and the beast says, ***okay,*** you have an half an hour to go / and then you will be back, or/ I will die. (Ar-Children, 1012)

In the BNC2014 children subcorpus, *okay* works as a filler to mark similar functions as in the Ar-Children corpus. However, I observed that sometimes speakers employ *okay* to mark their departure from the conversational setting, as in (8).

(8) S0416: may I please leave the table?

S0417: yeah

S0416: ***okay*** bye (SBCL / BNC2014 children)

In sum, speakers in both corpora employ DM *okay* to serve various textual functions. The filler function was more prevalent in the Ar-Children corpus compared to its use in the BNC2014 children subcorpus. I will discuss this results in more details in section 7.3.

6.2.4. Qualitative analysis of interpersonal functions of *okay* in context

Okay is predominantly an interpersonal marker. Its interpersonal functions include use as a turn management marker, a confirmation marker, a speech act marker and as a backchannel.

6.2.4.1. *Okay* as a turn management marker

This function is one of the most frequent functions of *okay* in both corpora, accounting for 96 instances of *okay* in the Ar-Children corpus (0.9 per 100 clauses) and 109 in the BNC2014 children subcorpus (0.5 per 100 clauses). As a turn management marker, *okay* expresses the speaker's desire to continue with their previous turn. For instance, in (9), speaker 121401 employs *okay* to link her previous turn and elaborate on the topic she has chosen, i.e. what type of job she would like to do. *Okay* displays her need to continue her own turn. This contribution shows that speakers in the Ar-Children corpus show awareness of their friends' ability (or sometimes lack of ability) to communicate effectively about certain topics. Determining this often required tracking back across longer stretches of the conversation.

(9) 121401: oh, job,

121402: yeah.

121401: ***okay***, like when , I want to be/ when I will, be bigger , I want to be/ like, mm doctor or, doctor or I dunno like doctor or teacher, and you/ (1214/ Ar-Children)

Using the turn management marker *okay* can signal a speaker's agreement with another speaker during the transition of turns, particularly following the second speaker's request. In (10), *okay* flags a response to the other speaker's direct request.

(10) 124502: yeah. okay, so can you now , ask me the question/

124501: ***okay,*** do you like, playing games, in phone, like, playing mm games,

124502: in the phone/ (1245/ Ar-Children)

Occasionally, *okay* can mark the speaker's desire to perform an act either during the conversation or in the future (see Condon, 1986). This subfunction co-occurs with "I will/'ll".

For instance, the two speakers in (11) have been talking about a play date at the park. *Okay* displays speaker 110102's decision to refer the matter to her own mother to specify the time for their play date.

(11) 110102: ***okay,*** I'll ask my mum which day,

110101: and, I'm gonna ask my mum, too, (1101/ Ar-Children)

Speakers in the BNC2014 children subcorpus also employ turn management marker *okay* to express acquiescence to another speaker's suggestion or request. For instance, in example (12), speaker S416 employs *okay* in response to a sibling's direct instruction.

(12) S0417: >>stop doing that with your feet sit up properly sit up properly young man

S0416: ***okay*** middle-age woman (SHX2/ BNC2014 children).

6.2.4.2. *Okay* as a confirmation marker

This function displays the reported original meaning of *okay*, as derived from the abbreviation of the phrase "oll Korrect" i.e. "all correct" (Klien, 2014). Confirmation marker *okay* appears twice as frequently in the Ar-Children corpus (0.9 per 100 clauses) as in the BNC2014 children subcorpus (0.4 per 100 clauses). *Okay* as confirmation marker expresses a speaker's understanding of other speakers. This includes, more precisely, *okay* signalling receipt of information (in second pair of a question-answer adjacency pair, see Sacks et al., 1974) or reply to answers (in a third pair in a sequence of a question-answer-reply) stated by other speakers.

Confirmation marker *okay* often appears in turn-final positions with a level or falling tone. This function appears 23 times in the Ar-Children corpus (0.2 per 100 clauses). In (13), speaker 080202 uses *okay* in turn final position to confirm her understanding of 080201's statement on colours.

(13) 080201: I like black, white/ and maybe red and pink.

080202: so mainly shades, ***okay,*** (0802/ Ar-Children)

There are 22 instances (0.2 per 100 clauses) of *okay* as a confirmation marker in the Ar-Children corpus that flag a speaker's attempt to confirm their understanding of another speaker's explanation. This function often appears after DM *oh* (see section 6.5.2.3) or with repetition of *okay*. In (14), Speaker 091601 employs *okay* twice to confirm that she recognises a mistake and understands the correction pointed out by the other speaker.

(14) 091602: grade, you are not six , يا ماما. (yā māmā 'oh mum': expression used in fighting mostly between friends)

091601: ***okay.*** ah grade two. ***okay***, sorry. (0916/ Ar-Children)

Speakers can also utilise confirmation marker *okay* to confirm that they have received an answer to their own earlier question (see section 6.2.2). This subfunction is prevalent in the Ar-Children corpus, accounting for 48 instances of *okay* as a DM (49.5%). *Okay* in this function, exemplified in (15), is the third part of a sequence where the first two parts form an adjacency pair, for an overall question/answer/response pattern (see Schegloff et al., 1977).

(15) 101501: you hate Corona/ or love Corona,

101502: I hate Corona,

101501: ***okay.*** like every person/ (Ar-Children, 1015)

6.2.4.3. *Okay* as a speech act marker

As a speech act marker, *okay* is more frequent in the Ar-Children corpus (0.4 per 100 clauses) than in the BNC2014 children subcorpus (0.1 per 100 clauses). Speakers employ speech act marker *okay* to seek information from other speakers, to request confirmation of others' understanding, or to initiate a direct or indirect request.

Speech act marker *okay* is sometimes followed by questions. Speakers employ the speech marker *okay* to seek more information in 9 examples in the Ar-Children corpus, including the speaker in (16), whose *okay* occurs in a rising tone unit prior to a question.

(16) 101201: uh first thing / actually, her mum. uh is- uh is dead. and her dad , also dead / so, she has mm a maid / a maid , and they- and they help her .

101202: ***okay/*** and what's happened to her, (1012/ Ar-Children)

Speech act marker *okay* can signal a confirmation check question. When used this way in turn-final position, speech act marker *okay* with rising tone appears to form a type of tag question (see Beach, 1993; 2020). In the Ar-Children corpus, speakers use *okay* to mark a confirmation check in 28 instances (0.4 per 100 clauses). In (17), Speaker 392901 employs *okay* to confirm that her friend is following her story. *Okay* with this function is often followed by an affirmative response by the other speaker, such as *uhu*, *mm*, *yes* or *yeah*.

(17) 392901: actually, do you know something. a story that's happened to me/

392902: what

392901: before I came to the school. I was in [ANON-place], ***okay/***

392902: uhu (3929/ Ar-Children)

In the Ar-Children corpus, there are also examples of speech act marker *okay* followed by a request, as in (18). In this example, speaker 110701 uses *okay* to flag her request to change the subject of discussion, which was about noodles called اندومي (ạndwmy, 'Indomie': a brand of instant noodles).

(18) 110702: actually , what I like to buy/ in the supermarket .اندومي, it's so yummy/

110701: yeah, I like it. but, not so much/ because, sometimes I feel it's not so yummy/ sometimes, I feel it, so delicious. and sometimes, I feel it, not delicious.

not yummy/ like in the middle,

110702: ***okay,*** let's talk about something else. (1107/ Ar-Children)

6.2.4.4. *Okay* as a backchannel marker

As a backchannel, speakers use *okay* not to take the floor or change the topic, but rather to confirm to another speaker that "I am happy for you to keep talking". Through this function, "[a] speaker remains speaker, and hearer remains hearer" (Schiffrin, 1987: 99). The backchannel function is one of the most frequent functions of *okay* in both corpora, accounting for 57 instances (0.7 per 100 clauses) in the Ar-Children corpus and 119 in the BNC2014 children subcorpus (0.5 per 100 clauses). Backchannel *okay* signals that speakers are requesting a continuation of their turn from other speakers.

In the Ar-Children corpus, backchannel *okay* often carries a level or falling tone. For instance, in example (19), speaker 122001 employs backchannel *okay* in a series of three turns with level or falling tone.

(19) 122002: I have two friend, you / and [ANON-name], but her name / is [ANON-name], is talking Urdu / Pashto/ yes / Pashto and she talk English.

122001: ***okay,***

122002: I go to her birthday in her home […]

122001: ***okay,***

122002: to play with it,

122001: ***okay.***

122002: she love because Barbie's/ […] (1220/ Ar-Children)

In the BNC2014 children subcorpus, speakers also employ backchannel *okay*, as in (20).

(20) S0415: but I miss you --ANONnameM () do n't go

S0416: ***okay***

S0415: do n't come back cos I do n't miss you

S0416: ***okay***

S0415: I 'm just teasing you --UNCLEARWOR

S0416: ***okay*** (SPC7/ BNC2014 children).

In summary, these four interpersonal functions are collectively the predominant uses of *okay* in both corpora. This result is consistent with previous studies that reported use of *okay* as a marker to engage other speakers. However, there are some differences in the frequencies of subfunctions between the L1 and L2 children such as the use of filler functions. I will discuss this further in chapter 7.

6.3. *Oh*

6.3.1. *Oh* in previous literature

Many prior studies have looked at DM *oh* as mainly serving interpersonal functions. One of the most studied uses of *oh* as interpersonal marker is marking a change-of-state in a speaker's turn (e.g. Heritage, 1984, 1998, 2018; Fox Tree & Schrock, 1999). In this function, *oh* can indicate either a change in the speaker's mind (resulting from being informed by or aware of prior talk) or express a change in emotional state. In the latter subfunction, *oh* often occurs adjacent to expressions of emotion such as pain, excitement or disappointment (Biber et al., 1999: 1083). James (1978: 518) differentiates between *oh* as a marker of emotion or realization, as in "*oh*, the house is burning down", and *oh* as a marker that flags the choice of a word from a set of alternatives as in "I bought, *oh*, a book". DM *oh* is also associated with the speaker's recognition of the present moment; for instance, Schouroup (1985: 13) describes *oh* as an *evincive* marker, that is, a marker that is "tied with the present moment of utterance" (Schouroup 1985: 14). Evincive markers signal real-time thinking and as such, aid speakers

in establishing the timeline of their thoughts with the real-time focus of a hidden mental event that might be brought into the shared conversation at a later point (Schouroup 1985: 15).

Heritage (1984) tries to combine both accounts of *oh* i.e. as an emotional marker and in the present moment marker, as part of marking a change-of-state. According to Heritage (1984: 302), *oh* can work as responsive marker to prior discourse to mark the informing function. As an informing marker, Heritage (1984: 301) argues that *oh* signals the "receipt" of information. Thus, speakers use *oh* to confirm that in the present moment, they are informed on the matter under discussion. In question-answer sequences, *oh* as a third turn by the questioner can flag the preceding answer as complete or informative. *Oh* can also function as a counter-informing marker, i.e. one that indicates a speaker is revising their previous position as a reaction to other speakers' contributions (Heritage, 1984: 312). *Oh* can also flag a speaker's acceptance of repairs initiated by other speakers. Heritage (1984: 318) argues that in this usage, *oh* illustrates speaker's understanding of other speakers' proposal of repair-initiation (see also Schegloff et al., 1977). Heritage's account of *oh* shows that *oh* works as a responsive marker that flags speakers' responses to other contributions.

Heritage (2018) also observes that the functions of *oh* can be interpreted according to its placement in conversational turn sequences. In a first-turn sequence (see section 6.2.4.1), *oh* has a verbal noticing function and marks the introduction of a new topic or information. *Oh* also indicates a change in speaker's state of knowledge to being informed by the answer of their questions in third turn sequence (Heritage, 1984). In second-sequence position, *oh* signals a response to an informing turn (as discussed earlier) or a question. Heritage (1998) argues that *oh* can precede question-answers to indicate that the question is problematic, with regard to its relevance or presupposition. Thus, *oh* expresses that the question has prompted "a shift in the respondent's attention" (Heritage, 1998: 326).

Schiffrin (1987) agrees on the previously discussed pragmatic function of *oh* as an indicator of a shift of speaker's orientation to information. She also argues that *oh* can also carry an interactional function – similar to an interpersonal function in my account. The interactional function of *oh* occurs in its use as a response to unexpected questions or questions that request clarification. When *oh* marks a response to a question, it confirms the speaker's engagement in the conversation as an active receiver of information. With this function, *oh* provides time for speakers to process the information while displaying their presence in the conversation (Schiffrin, 1987: 100).

Fox Tree & Schrock (1999) examine listeners' comprehension of words preceded by interpersonal marker *oh*, in comparison to the same material without *oh* or with *oh* replaced by a pause. They find that listeners recognize words faster if preceded by *oh*. According to Fox Tree & Schrock (1999: 290), this confirms the hypothesis of *oh* as a change-of-state marker, since it prompts the other speakers to focus on the relation between the upcoming discourse with prior talk. Through this use, *oh* increases the other speakers' attention to previous information at the point of relevance. *Oh* can co-occur with repairs and as such, it can assists integration in speakers' management of discourse by wiping out confusion that may arise if a speaker stops their projected utterance and launches a new one without *oh* (Fox Tree & Schrock, 1999: 294).

Aijmer (2002:99) differentiates between *oh* uses as a backward-looking and as a forward-looking marker. As a backward-looking marker, *oh* carries functions such as organising the discourse, or introducing a correction or an explanation (akin to the textual functions in my framework). As a forward-looking marker, *oh* is used as intensifying marker (similar to the interpersonal functions in my framework). Aijmer (2002: 151) argues that use of *oh* does not simply signal a change-of-state but rather indicates interactional functions that depend on its context. Bolden (2006: 677) similarly claims that *oh* reveals more than the

speaker's state of mind. For instance, *oh* can mark an utterance as "being just now remembered". *Oh* in this function does not flag speaker's internal state, but rather offers an account of how a speaker chooses to initiate their upcoming utterance at a specific moment in the conversation (Bolden, 2006: 673).

Oh has rarely been studied in L2 speakers' discourse. Huang (2011) investigates the use of *oh* in the corpus of Chinese speakers of L2 English in comparison to reference corpora of American English and British English speakers as well as contrasting dialogic and monologic genres. Huang (2011: 104) notes that there is no difference in the frequency of DM *oh* between L1 and L2 speakers. However, *oh* tends to be significantly more frequent in dialogic contexts than in monologues. Huang finds that *oh* marks various discourse functions, including quotations, hesitations, repairs, and responses to a question. Additionally, it is employed to open or close a topic, to show emotion, and to signal completion of a cognitive process. Chinese speakers are observed as tending to use *oh* as an acknowledgment marker – a function not found in the L1 corpora.

In sum, *oh* has been studied as, mainly, a marker of change-of-state. However, *oh* can signal a range of interpersonal and textual functions.

6.3.2. Frequencies of the functions of DM *oh*

The overall frequency of *oh* as a DM and non-DM is presented in table 6.3. The non-DM use of *oh* was discussed earlier (section 3.5.1).

Table 6.3. *Raw & relative frequencies (per 100 clauses) of* oh *in the Ar-Children corpus and the BNC2014 children subcorpus*

Corpus	Total use of *oh*	DM use	Non-DM use
Ar-Children	317 (3.0)	294 (2.8)	23 (0.2)
BNC2014 children	1,068 (4.7)	1,029 (4.5)	39 (0.2)

After excluding non-DM instances of *oh*, I applied the framework for DM functions (see section 3.6) to the data. Table 6.4 presents the frequencies of each function of *oh* in the Ar-Children corpus and in the BNC2014 children subcorpus.

Table 6.4. *Raw and relative (per 100 clauses) frequencies and percentages of* oh *with different functions in the two corpora*

Function		**Ar-Children**			**BNC2014 children**		
		N	Per 100	%	N	Per 100	%
Textual	initiator	26	0.2	8.8	131	0.6	12.7
	elaborator	6	0.1	2.0	38	0.2	3.7
	filler	23	0.2	7.8	61	0.3	5.9
	transitional	28	0.3	9.5	108	0.5	10.5
	Total	**83**	**0.8**	**28.2**	**338**	**1.6**	**32.8**
Interpersonal	speech act	13	0.1	4.4	36	0.2	3.5
	responsive	64	0.6	21.8	179	0.8	16.4
	understanding	78	0.7	26.5	189	0.8	18.4
	emotion	39	0.4	13.3	64	0.3	6.2
	backchannel	17	0.2	5.8	121	0.5	11.8
	Total	**211**	**2.0**	**71.8**	**589**	**2.6**	**56.3**
Unclear		0	0	0.0	102	0.5	9.9
Total		**294**	**2.8**	**100**	**1,029**	**4.7**	**100**

The results show that *oh* is predominantly an interpersonal marker in both corpora. It is frequently used with responsive function and as a marker of understanding and emotion, among other functions.

However, speakers in the BNC2014 children subcorpus employ *oh* more frequently as a textual marker than speakers in the Ar-Children corpus. In particular, *oh* appears more frequently as a transitional marker and initiator in the BNC2014 children subcorpus.

The differences across the two corpora between the frequencies of the functions of *oh* were tested for significance using the Log-Likelihood test. To perform the test, I followed the same steps as used for other DMs (see section 5.2.2). The functions of *oh* differ significantly between the two corpora at the $p<0.001$ level (LL= 146.03, df = 10).

6.3.3. Qualitative analysis of textual functions of *oh* in context

Oh appears as a textual marker in 83 examples in the Ar-Children corpus and 338 in the BNC2014 children subcorpus. These examples are spread over the four main textual functions: initiator, elaborator, transitional and filler.

6.3.3.1. Initiator *oh*

The initiator *oh* occurs three times more frequently in the BNC2014 children subcorpus (0.6 per 100 clauses) than in the Ar-Children corpus (0.2 per 100 clauses). The initiator function of DMs is the use of a DM to introduce a new topic or piece of information (see section 3.6). Initiator *oh* is used mainly to introduce new information in both corpora, as illustrated in (20).

(21) 120801: I 'm so slow. I 'm so slow

120802: now movies, ***oh***, I watched to- to a movie/ no way Home, in the break.

The introduction of new information is sometimes triggered by something said by other speakers, as is indeed the case in (22), where the novel information after *oh* is parallel to the information in the preceding speaker's turn (both concern family members' school grades). In the Ar-Children corpus, these kinds of examples of *oh* often carry level tone – either by itself or as part of a longer prosodic unit.

(22) 693002: yeah. my sister. she's in grade six,

693001: ***oh*** I have like, a small sister in KG. me and like, my big brother/ I think he's in grade nine or ten, (6930/ Ar-Children)

6.3.3.2. Elaborator *oh*

Elaborator *oh* is one of the least commonly used functions of DM *oh*, with 6 instances in Ar-Children corpus (0.1 per 100 clauses) and 38 instances in the BNC2014 children subcorpus (0.2 per 100 clauses). Elaborator *oh* signals that the following utterance is an explanation for a prior utterance, often within the same turn, as in (23), where speaker 094101 uses *oh* prior to explaining the meaning of "army and Blink".

(23) 094101: No. I 'm army and Blink, I do n't know. ***oh*** that's- I am K-pop, I love Kpop/ (0941/ Ar-Children)

Because it often flags elaboration of an utterance earlier in the turn, in both corpora elaborator *oh* typically appears in turn-medial position.

6.3.3.3. Filler *oh*

In both corpora, filler *oh* flags a self-repair or planning functions. This function is infrequent in both corpora, accounting for 23 instances of *oh* in the Ar-Children corpus (0.2 per 100 clauses) and 61 in the BNC2014 children subcorpus (0.3 per 100 clauses).

As a self-repair marker, *oh* often appears in turn-medial position (Aijmer, 2002: 99), to indicate that the following utterance is a repair to the prior one. In (24), for instance, speaker 092902 uses *oh* before an immediate adjustment to the content of the preceding utterance.

(24) 092902: mm I kind of like math but sometimes it's really ugly and hard, and yeah / science is also cool. I do n't really have a favourite subject, ***oh***/ yeah, mm PE, uh and art, (0929/ Ar-Children)

In the Ar-Children corpus, filler *oh* is more commonly used (along with a filled or unfilled pause) as a planning marker, accounting for 17 instances (0.2 per 1000) as in (25).

(25) 104801: mm I do n't know, I think/ yes, but/ I do n't know, I 'm gonna ask the teacher/ before I go/ () out the school/ I'm gonna ask her, uh what should we study for
104802: mm ***oh***. okay/ (1048/ Ar-Children)

In the BNC2014 children subcorpus, filler *oh* sometimes appears in turn-medial position to flag the speaker's self-repair of a prior utterance, as in (26).

(26) S0416: I do n't really remember anything else though ***oh*** I do I do remember a bit yeah that was okay that was okay (S3CK/ BNC2014).

6.3.3.4. Transitional *oh*

Transitional *oh* marks a change of topic (in which case it is often followed by a question) or a shift to direct speech. It also signals sequential relations between utterances. As a transitional marker, *oh* occurs more frequently in the BNC2014 children corpus (0.5 per 100 clauses) than in the Ar-Children corpus (0.3 per 100 clauses). Example (27) illustrates the transitional *oh* indicating change of topic, with a following question.

(27) 120802: I eat chips, all in one day.
120801: ***oh*** what's your favourite animal /
120802: animals it's rabbit. it's so cute, (1208/ Ar-Children)

The other use of transitional *oh* is as a quotative signal, marking boundaries between the speaker's present utterance and the reported speech (Biber et al, 1999: 1118). Quotative *oh* is one of the less frequently occurring functions in the Ar-Children corpus, since there is, overall, very little reported speech in this data. Most occurrences of quotative *oh* are preceded

by quotative *be+like,* as in (28); other examples are prefaced with reporting verbs such as *tell* and *say*.

(28) 835001: you just, like, text like, so, I can see. if-if I 'm wrong or no , like , something wrong [ANON-name] , and then , I'm just , like , ***oh*****/** I 'm sleeping , what the heck , you-you tickle me every day. (8350/ Ar-Children)

Quotative *oh* is more prevalent in the BNC2014 children subcorpus, accounting for half the instances of transitional *oh* (57 examples, 30 with *be+like*). Example (29) shows use of quotative *oh* after the reporting verb *said*.

(29) S0622: well it's what I've said to them if him er he said ***oh*** I 'm not going there again good

S0525: oh he said he was n't gonna go again? (SR7C/ BNC2014)

Transitional *oh* can also mark sequential relations between utterances when it occurs in turn-medial position. In this case, the sequential marker *oh* signals a change of perspective on part of the speaker (Aijmer, 2002: 99) and therefore acts as a link between two utterances within the same turn. In other words, it signals that the idea has just arrived in the speaker's mind while they were talking. In (30), the speaker moves from searching for a topic of discussion to stating the topic that she comes up with (i.e., going to a shopping mall).

(30) 94101: yeah. okay, so. mm what were we talking about, ***oh.*** what we go, like, shopping with [ANON-name], and you/ and me, all three. (9410/ Ar-Children)

In sum, speakers in both corpora are able to employ *oh* with the four main textual functions, albeit the frequencies are low.

6.3.4. Qualitative analysis of interpersonal functions of *oh* in context

6.3.4.1. *Oh* as a speech act marker

DM *oh* can be said to operate as a speech act marker when it occurs in elicitation turns to keep the conversation going (Aijmer, 2002: 123). Questions prefaced with *oh* are often projected towards previous information. Such questions are used to request either a clarification or a confirmation of given information. Such questions are called elaboration questions (Schiffrin 1987: 85) or reaction questions (Aijmer, 2002: 124). The speech act marker function is infrequent in both corpora compared to other functions of *oh*, accounting for 13 instances in Ar-Children corpus (0.1 per 100 clauses) and 36 in the BNC2014 children subcorpus (0.2 per 100 clauses). One instance is given in (31), where *oh* flags the speaker's expression of interest in finding out more about her friend's holiday.

(31) 083402: uh طيب . (*tayeb* 'good, okay, fine') who- where do you like to go in the holiday,

083401: I like-I like to go to my, with my family /

083402: ***oh***, to where, (0834/ Ar-Children)

In (32), the question prefaced by speech act marker *oh* works to elicit confirmation of an assumption made by the speaker about a preceding narrative. That narrative constitutes contextual information that was recently provided and which triggers a memory in the speaker's mind.

(32) 094601: I draw a very nice , like , flower /

094602: *well* that's cool /

094601: yeah.

094602: ***oh*** do you mean the one that you, you showed our teacher, (0946/ Ar-Children)

6.3.4.2. *Oh* as a responsive marker

Oh can appear in turn-initial position to flag the utterance as a response to a question or to prior uttered information. Heritage (1984: 310) notes that speakers use *oh* in this way to show that they are now informed by the new information. This responsive marker function is among the most frequent functions in both corpora. As a responsive marker to questions, *oh* signals that a response implies that either the question is problematic or the answer may not be explicit. In the Ar-Children corpus, responsive marker *oh* accounts for 64 occurrences of DM *oh* (0.6 per 100 clauses), of which 21 are of the type marking a response to a question. This is in line with Huang's finding that L2 learners tend to use *oh* as a preface to a response to a question (Huang, 2011: 130). In the Ar-Children corpus, speakers sometimes use responsive *oh* with a rising tone to express their surprise at the question, as seen in (33).

(33) 084602: so. mm, () who's your favourite teacher/

084601: ***oh*****/** this is a nice one, I 'm not choosing favourites, I like all of them/ (0846/ Ar-Children)

By contrast, in (34), *oh* expresses that the answer is not only that a speaker wishes for a cat but continues to introduce the fact that she already has a hamster.

(34) 110101: […] do you like dogs/ or cats more,

110102: ***oh***, I wish, I had a cat, but I have a hamster/ (1101/ Ar-Children).

Oh as a responsive marker can signal that the speaker has just remembered some information relevant to the conversation topic, that is, it is a case of the use of a DM accompanying, and signalling, some cognitive process (Bolden, 2009: 673, see also Heritage, 1984, 1998). For instance, in (35) speaker 103701 initially seems to have forgotten about a

math exam; the use of responsive marker *oh* (with a rising tone) is triggered by the cognitive process whereby the speaker remembers. Furthermore, *yes* confirms that what has been remembered is in line with the other speaker's assertion about the exam.

(35) 103702: […] how did you do on the math exam,

103701: I did- we never had the math exam,

103702: we did, online,

103701: ***oh*****/** yes, the math exam, uh it was very scary. (1037/ Ar-Children)

6.3.4.3. *Oh* as a marker of understanding

Oh-prefaced utterances can express understanding of prior information or agreement with other speakers' views. This function as a marker of understanding is one of the most prevalent functions of DM *oh* in both corpora, accounting for 78 instances in the Ar-Children corpus (0.7 per 100 clauses) and 189 in the BNC2014 children subcorpus (0.8 per 100 clauses). These examples of *oh* often co-occur with other DMs, such as *yeah*, *yes* and *okay*. The co-occurrence can make the response seem politer to other speakers (Fox Tree & Schrock, 1999: 282). In (36), speaker 24502 expresses her understanding of her friend's preceding utterance through the words *I understand*, but further reinforces this assertion with the use of *oh* (alongside DM *yeah*).

(36) 124502: mm uh do you like/ to spend your time, with your family/ of your- or your- or with your friends.

24501: uh with my friends. like, uh they uh make, uh like fun uh things , we play together,

24502: ***oh***, yeah. I understand, (245/ Ar-Children).

Similarly, *oh* can reinforce agreement if accompanied by other agreement DMs, or if it prefaces entire agreement utterances, as in (37).

(37) 090601: […] what about- what 's your like favourite school subject /

090602: probably science.

090601: mine is probably English / cause it's the easiest,

090602: ***oh*** yeah, I agree. (0906/ BNC2014 children)

Oh can additionally be used alone, as a full, free-standing turn to show understanding of a prior question-answer pair. Heritage (1984: 307) calls *oh* with this function a "receipt of information". In other words, *oh* signals that the speaker's state of mind has changed from uninformed to informed (Schiffrin, 1987: 115). Question-askers reply to answers with *oh* to declare their understanding, and to alert the answerers that the answer has been sufficient or appropriate to their question (Fox Tree & Schrock, 1999: 282). In (38), speaker 103701 uses *oh* to express her understanding of her friend's answer to her question. *Oh* in this example has rising tone, which supports this interpretation.

(38) 103701: what's a PS four,

103702: Play Station.

103701: ***oh***/

103702: yeah, (1037/ Ar-Children)

6.3.4.4. *Oh* as an emotion marker

Oh can be used to express emotion such as disappointment, surprise or excitement (James, 1974: 518). It is also used to emphasise positive adjectives or adverbs (Aijmer, 2002: 101). In the Ar-Children corpus, *oh* intensifies adjectives like *good*, *nice* and *cool*. Additionally, it accompanies other words such as *wow* to show surprise and *no* and *sorry* to show empathy, as in (39).

(39) 093001: I also forgot to do my homework.

093002: ***oh*** no, that's sad, (0930/ Ar-Children)

Oh as a marker of emotion appears in 39 examples (0.4 per 100 clauses) in the Ar-Children corpus and in 64 examples (0.3 per 100 clauses) in the BNC2014 children subcorpus. An example is (40), where speaker S0655 uses *oh* to express disappointment about having a clown in the festival.

(40) S0653: mm well yes that is exaggeration but I expect there will be clowns

S0655: ***oh*** no

S0653: I know but hopefully

S0655: clowns are scary (SP5V/ BNC2014).

The emotion marker *oh* signals surprise, disappointment and excitement in the BNC2014 children subcorpus. It is also used as an intensifier with words like *yay*, *excellent* and *thanks*.

6.3.4.5. *Oh* as a backchannel marker

Oh as a backchannel marker occurs as a standalone turn. It does not mark a change of turn, since "speaker remains speaker and hearer remains hearer" (Schiffrin 1987: 99). Furthermore, use of *oh* as a backchannel marker does not open a new topic or shift an ongoing topic (Aijmer, 2002: 135). Its purpose in this function is to show the speaker's engagement and thus, to encourage the other speaker to continue (see section 6.2.4.4). In the Ar-Children corpus, *oh* as a backchannel marker is less frequent than in the BNC2014 children subcorpus, accounting for 17 (0.2 per 100 clauses) examples in the Ar-Children corpus, compared with 121 (0.5 per 100 clauses) examples in BNC2014 children subcorpus. Backchannel marker *oh* often carries a level tone, as in (41).

(41) 110202: to be honest like , I've painted a lot of stuff that is not my favourite , like but my favourite painting that I have/ it 's like this . () house / it 's like a garden, you know/

110201: ***oh,***

110202: I tried to make it realistic, I think you saw it/ in grade four/ did you see it/

(1102, Ar-Children)

In summary, *oh* is predominantly an interpersonal marker. In both corpora, speakers employ it to show their engagement in the conversation by responding to other speakers' turns. In addition, they use *oh* to mark request for clarification, confirmation or to continuation.

6.4. *Yeah*

6.4.1 *Yeah* in previous literature

Yeah is known as a minimal response marker that mainly expresses agreement. Studies have shown that *yeah* is often an interpersonal marker that can indicate speaker engagement in a conversation (e.g. Schegloff, 1982; Jefferson, 1983; Drummond & Hopper, 1993a, 1993b, 1993c; Gardner, 1998; Fuller, 2003b; House, 2013; Haselow, 2017; Thuy, 2019). Drummond and Hopper (1993a) argue that *yeah* can also carry various discourse functions. Jefferson (1983: 5-6) argues that *yeah* can express an intention to take over the conversational floor or to change the conversational topic (see also Jucker & Smith, 1998: 181). Speakers can also employ *yeah* in turn-initial position to flag an elaboration on another's turn (Drummond and Hopper, 1993a: 53). Schegloff (1982: 78) argues that *yeah* can also signal a speaker's attempt to discourage other speakers from initiating a repair to their previous turn. In this usage, *yeah* aids in mitigating any potential misunderstanding on the part of the other interlocutors (Haselow, 2017: 162).

The functions of *yeah* differ according to context and speakers. For instance, Fuller (2003b) observes that *yeah* is less common in conversation than in interviews (see also Liao,

2009). This could be because its use is influenced by the type of relationship between interlocutors. Jucker and Smith (1998) find that *yeah* is more common between strangers, which explain its observed prevalence in interviews in other studies. In addition, researchers have found that *yeah* is more frequent in the speech of L2 speakers than L1. For instance, using data from prior studies, Gardner (1998) finds that except for *yeah*, L2 speakers employ acknowledgment markers including *oh, mm, mm hmm, right* and *okay* less often than do L1 speakers.

Gardner (1998) subsequently undertakes an analysis of discourse functions of *yeah* in a corpus of American, Australian and British English speech. Gardner finds that *yeah* has two main discourse functions: as a continuer of a previous speaker's turns to indicate strong alignment, and as an acknowledgment marker. Speakers also employ *yeah* to encourage other speakers to continue. In this function, *yeah* carries a rising intonation contour to express speaker's involvement with another speaker's narrative (Gardner, 1998: 219). Additionally, speakers employ *yeah* to show collaboration in a conversation through participating in another speaker's narrative (Gardner, 1998: 216).

In a study of *yeah* in English telephone conversations by L1 and L2 speakers, Wong (2000: 61) finds that L1 speakers rarely employ *yeah*. When they do, they use it in turn-initial position, mainly to inform other speakers of upcoming repair. In contrast, Chinese L2 speakers use *yeah* to signal different self-initiated repair techniques in turn-medial position. For these speakers, *yeah* can alternatively signal the announcement of a search for words, continuation or the end of a search. Thus, Wong finds that in these L2 learners' speech, *yeah* is mainly a filler that signals repair or disfluency. Bu (2013) finds that Chinese learners of English employ *yeah* with various conversational functions, including turn transition, confirming another's turn and elaborating or commenting on previous turns. Looking at German L2 learners, House (2013) finds that *yeah* works as a marker to organise their

contributions to the conversation and as a marker of acknowledgement of the contributions of previous speakers.

In summary, *yeah* can mark various discourse functions in the speech of L1 and L2 speakers of English. The different accounts in the literature of use of *yeah* by L2 speakers highlight the need to study this marker in different contexts, including how it is used by different groups of speakers.

6.4.2. Frequencies of the functions of DM *yeah*

Yeah is one of the most frequent words in both corpora, ranked 17th in the Ar-Children corpus and 8th in the BNC2014 children subcorpus (see Appendix F). The frequencies of *yeah* as a DM and non-DM are presented in table 6.5.

Table 6.5. *Raw & relative frequencies (per 100 clauses) of* yeah *in the Ar-Children corpus and the BNC2014 children subcorpus*

Corpus	Total use of *yeah*	DM use	Non-DM use
Ar-Children	848 (7.9)	639 (6.0)	209 (1.9)
BNC2014 children	2,653 (10.7)	1,668 (7.4)	985 (3.3)

Table 6.6 presents the different functions of *yeah* as a DM in the two function domains of textual and interpersonal.

Table 6.6. *Raw and relative (per 100 clauses) frequencies and percentages of* yeah *with different functions in the two corpora*

Function		**Ar-Children**			**BNC2014 children**		
		N	Per 100	%	N	Per 100	%
Textual	elaborator	15	0.1	2.3	45	0.2	2.7
	filler	19	0.2	3.0	45	0.2	2.7
	transitional	38	0.4	5.9	38	0.2	2.2
	Total	**72**	**0.7**	**11.3**	**128**	**0.6**	**6.6**
Interpersonal	turn manage.	159	1.5	24.9	325	1.4	19.5
	confirmation	127	1.2	19.9	185	0.8	11.1
	agreement	88	0.8	13.8	235	1.0	14.1
	speech act	15	0.1	2.3	20	0.1	1.2
	attitude	19	0.2	3.0	31	0.1	1.9
	backchannel	159	1.5	24.9	644	2.8	38.6
	Total	**567**	**5.3**	**88.7**	**1,440**	**6.4**	**86.4**
Unclear		0	0	0	100	0.4	6.0
total		**639**	**6.0**	**100**	**1,668**	**7.4**	**100**

As a DM, *yeah* marks interpersonal functions more than 80% of the time in both corpora. The relative frequencies of *yeah* as turn management marker and as backchannel in the Ar-Children corpus are similar. The backchannel function of *yeah* is almost twice as frequent in the BNC2014 children subcorpus (2.8 per 100 clauses) as in the Ar-Children corpus (1.5 per 100 clauses). *Yeah* functioning as a confirmation marker is the third most frequent function in both corpora. *Yeah*'s textual DM subfunctions (transitional marker,

elaborator, filler) have similar relative frequencies in the BNC 2014 children subcorpus, but in the Ar-Children corpus *yeah* is more frequently a transitional marker.

The differences across the two corpora between the frequencies of the functions of *yeah* were tested for significance using the Log-Likelihood test. The functions of *yeah* differ significantly between the two corpora at the $p<0.001$ level (LL= 155.88, df = 10).

The next section will elaborate on the functions of *yeah* as a DM in detail, providing examples from both corpora.

6.4.3. Qualitative analysis of textual functions of *yeah* in context

Instances of *yeah* serving functions in the textual domain signal organisation and structuring of discourse. There are three main functions of *yeah* in the textual domain, namely elaborator, filler and transitional functions.

6.4.3.1. Elaborator *yeah*

As an elaborator, *yeah* introduces additional information to clarify a prior adjacent utterance within the same turn. This function is infrequent in both corpora. Speakers in the Ar-Children corpus use the elaborator *yeah* 15 times (0.1 per 100 clauses). When used to mark explanations, *yeah* mainly occurs in turn-medial position. In (42), where the context is a discussion of kinds of noodles, speaker 094101 wants to reinforce the point with an explanation. *Yeah* acts as a link to the second utterance in which the explanation – that the Korean type of noodles اندومي (ạndwmy, 'Indomie': a brand of instant noodles) has a pink container – is provided.

(42) 094101: yeah I know, اندومي in the Korean. Korean. ***yeah,*** the pink one, I really want, like, to eat it/ because I never eat it/ like, you know. (0941/ Ar-Children)

In the BNC2014 children subcorpus, there are 45 instances of *yeah* (0.2 per 100 clauses) as elaborator. In addition, *yeah* marks a speaker's attempt to elaborate on their previous utterances through providing examples. For instance, in (43), after answering their mother's question, speaker S0655 employs *yeah* in turn-medial position to offer an exemplification.

(43) S0654: there was one where erm like every single child's running towards – UNCLEARWORD

S0653: the camera?

S0655: yeah and going towards the camera like this no running towards the camera cos they can walk by then ***yeah*** it 's like starting from a baby and then getting older to where they are (SES2/ BNC2014 children)

6.4.3.2. Filler *yeah*

Yeah as a filler signals a speaker's attempt to repair their prior utterances or as a planning marker that often co-occurs with filled and unfilled pauses. Unlike the planning function identified for other DMs, *yeah* as a planning filler is said to display the completion of the planning process to affirm that the speaker has reached clarity in what they intend to say (see Wong, 2000: 61). The relative frequency of *yeah* as a planning filler is similar in the two corpora, accounting for 18 instances of *yeah* (0.2 per 100 clauses) in the Ar-Children corpus and 29 instances (0.2 per 100 clauses) in the BNC 2014 children subcorpus. In (44), speaker 090802 uses the filled pause (*uh*), and then stops to think for a short period (less than five seconds). The use of *yeah* indicates that she has finally identified what she was looking for.

(44) 090802: uh I like-like so much things,

090801: like what,

090802: like, uh () ***yeah,*** pasta. (0908/ Ar-Children)

Yeah as a planning marker exhibits similar functions in the BNC children subcorpus. For instance, one speaker in (45) employs *yeah* after a pause and hesitation (*erm*) while thinking about whether they know what the Kremlin is.

(45) S0514: it's like their Big Ben

S0512: >>Kremlin () no it erm ***yeah*** I have heard of it (SASC/ BNC2014 children)

The filler *yeah* can also signal a repair to a previous utterance. This function is rare in the Ar-Children corpus with only two examples, including (46), where *yeah* is followed by a correction of *I* to *we*.

(46) 094102: and I forget, ***yeah***, and we forget. uh what's happened I think/ yeah. he was shocked (0941/ Ar-Children)

The repair function appears in 16 instances of *yeah* (0.1 per 100 clauses) in the BNC2014 children subcorpus. In (47), for instance, *yeah* introduces speaker S0416's attempt to repair their previous utterance with a different grammatical structure.

(47) S0416: >>and like () he was like r- ***yeah*** like has like really healthy stuff () I mean like () yeah and it 's like and he does tae kwon do and stuff most weeks and he 's been doing it for years (SV5A/ BNC2014 children)

6.4.3.3. Transitional *yeah*

Speakers in both corpora use *yeah* to signal a transition in their speech, toward a closure or to change or resume topics. In the BNC2014 children subcorpus – but not the Ar-Children corpus – it is also seen to mark a transition to quoted speech. The transitional function is twice as frequent in the Ar-Children corpus (0.4 per 100 clauses) as in the BNC2014 children subcorpus (0.2 per 100 clauses).

In the Ar-Children corpus, *yeah* marks a transition to a closure in 23 instances (0.2 per 100 clauses). For instance, in (48) Speaker 092902 uses *yeah* to signal that her listing of items is complete.

(48) 092902: and I got, uh () what is it called anyway, some more clothes, and a water bottle, and ***yeah***. and the lunchbox , ***yeah/*** (0929/ Ar-Children)

Sometimes, the transitional *yeah* flags the end of a conversation. All the instances of *yeah* as a closure marker to end the conversation in the Ar-Children corpus co-occur with DM *okay* (see section 6.2.3.3). For instance, in (49) speaker 093002 employs *yeah* alongside *okay* to signal her desire to end the conversation.

(49) 093001: oh, me too I love computer, and science, sometimes science,
093002: ***yeah,*** okay, I have to go now, bye,
093001: bye, (0930/ Ar-Children)

When *yeah* signals transition to a new topic or resumption of an old one, the speakers create coherence by linking their turns with earlier speech. This function accounts for 15 examples of transitional *yeah* in the Ar-Children corpus. For instance, in (50), speaker 103702 uses *yeah* in turn-initial position to signal a change of topic from fishing to their last holiday trip.

(50) 103701: I was-I was/ trying to catch a fish, sin- since four years. and I-and I just only once, caught it ,
103702: yeah/ catching fish is hard,
103701: probably/
103702: ***yeah,*** I travelled/ to [ANON-place], two days ago, () it was an interesting trip. (1037/ Ar-Children)

The use of *yeah* to mark topic resumption is rare in the BNC2014 children subcorpus, with just two examples. In both cases, the speaker has been interrupted by another child, as in (51), where following the interruption by S0418, speaker S0416 uses *yeah* to announce a return to the earlier topic (comics).

(51) S0416: I think they 're called () they 're like comics

S0417: yeah

(..) S0418: mummy

[…]

S0417: *yeah* it is nice that he says our names now

S0416: *yeah* () ***yeah*** I read some graphic novels at school (SYTL / BNC2014 children)

As noted earlier, *yeah* signalling transition to quoted speech is found only in the BNC2014 children subcorpus (9 instances), as in (52).

(52) S0417: --ANONnameM you should have said and just sort of come off

S0416: I no I did I was I 'm fine like everyone was like oh you alright? you alright ? I was like ***yeah*** I'm fine (SX2V/ BNC2014 children)

In sum, *yeah* as a textual marker is infrequent in both corpora, but nevertheless speakers demonstrate capacity with different textual functions.

6.4.4. Qualitative analysis of interpersonal functions of *yeah* in context

Yeah fulfils different functions in the interpersonal domain, primarily aiding the continuity of conversation between interlocutors.

6.4.4.1. *Yeah* as a turn management marker

As a turn management marker, *yeah* expresses a speaker's desire to take over the conversational floor e.g., when interrupted, to continue their turn, or to stop other-initiated repairs (Schegloff et al., 1977: 367). The turn management function of *yeah* is the second most frequent of its functions in both corpora, accounting for 153 (1.5 per 100 clauses) examples of *yeah* in the Ar-Children corpus and 325 instances (1.4 per 100 clauses) in the BNC2014 children subcorpus. Turn management *yeah* often appears in turn-initial position, consistent with Haselow (2017: 5).

There are 26 instances in the Ar-Children corpus of turn management *yeah* operating as a floor-holding marker. *Yeah* with this subfunction is used by speakers as a way to hold the conversational floor in order to maintain the flow of a narrative, for example. For instance, in (53) speaker 103702 uses *yeah* to show her agreement with the other speaker's interjected comment, but also to declare her wish to hold the conversational floor and continue her narrative.

(53) 103702: […] so, mm he put a full / like duck, on like the stove / like not even with a pan. or anything , he just put it /
103701: mm that's how/ people the ancient people used to cook/ actually,
103702: ***yeah***, and he put it on, and like. it was n't even clean from the inside […] (1037/ Ar-Children)

A similar subfunction of *yeah* is its use as a marker to stop other-initiated repair (see Schegloff et al., 1977, Schegloff, 1982, Wong 2000). In (54), speaker 083402 interrupts her friend's story to assist her in finding the correct expression. *Yeah* expresses the speaker's 083401 desire to keep the conversational floor and continue her story.

(54) 083401: yeah. I do n't uh- I said-I said the- I do n't say the real name. I said, like I said

083402: like a fake name, yeah.

083401: ***yeah***, then he said mm come, I'm gonna give you a chocolate, and I said no, no, I do n't want. (0834/ Ar-Children)

Speakers often employ turn management *yeah* to elaborate on other speakers' comments or narratives. In (55), the use of turn-initial *yeah* signals speaker 122002's attempt to elaborate on the other speaker's turn.

(55) 122002: […] I love, mm clothes more, they have/ so much design in it.

122001: ***yeah***, and especially, like, and especially like. mm the last time I've gone shopping with my mum, and my two brothers/ mm we've gone to Max. and Centerpoint. and like I buyed from Max, this beautiful jacket it like. black jackets, (1220/Ar-Children)

Use of turn management *yeah* is similar in the BNC2014 children subcorpus. For instance, in (56), speaker S0655 employs *yeah* to hold the floor to continue a previous turn. This subfunction of *yeah* is twice as frequent in the BNC2014 children subcorpus (0.2 per 1000) as in the Ar-Children corpus (0.1 per 1000).

(56) S0654: to tell them I l- that sometimes I get really really upset and really really really erm

S0653: hot

S0654: yeah and really really hot if I ha- if I wear loads of layers

S0655: >>although

S0653: I will

S0655: ***yeah*** although they do make you wear lots of layers (S7RD/ BNC2014 children)

An example from the BNC Children subcorpus of the floor holding subfunction discussed above is presented in (57), as speaker S0655 employs *yeah* to keep the conversational floor after being interrupted by their mother's comment.

(57) S0655: you just speed off and sick six means that it 's quite easy

S0653: well --ANONnameM you do fine on the bike I have n't got any concerns

S0655: ***yeah*** you turn it to sixth gear when you're going downhill I think or something or maybe first gear (S839/ BNC2014 children)

An example from the BNC2014 children subcorpus of floor holding with *yeah* after an other-initiated repair is shown in (58). Here, speaker S0417 tries to offer the word that other speaker has been struggling to express. After saying *yeah*, S0416 acknowledges the suggestion but then continues to complete their turn.

(58) S0416: or it's like it's called like ravi- ravi- yes

S0417: >>ravioli

S0416: ***yeah*** something like that and it looks and tastes and smells disgusting in my opinion (S3CK/ BNC2014 children)

One of the subfunctions of turn management *yeah* that is prevalent in the BNC2014 children subcorpus is its use prior to elaboration on another speaker's narrative or comments. This function is more frequent here (1.4 per 100 clauses) than in the Ar-Children corpus (0.9 per 100 clauses). In (59), speaker S0653 is explaining the precise process of baking a chocolate cake to the children. Speaker S0655 uses *yeah* to interrupt her and add their own explanation.

(59) S0654: >>it will just and also if it's too hot it will become more runny and

S0653: no no it's not if if it's too hot the chocolate will separate which is disgusting it's not nice at all --UNCLEARWORD

S0655: ***yeah*** and then you get like really dark chocolate and really light chocolate (S527/ BNC2014 children)

6.4.4.2. *Yeah* as a confirmation marker

Speakers use *yeah* to confirm their own or others' utterances. When *yeah* as a confirmation marker appears in turn-final position, it strengthens the illocutionary force of the utterance, which Haselow (2017: 183) describes as "echoing reinforces". This function is the third-most prevalent function of *yeah* as an interpersonal marker in the Ar-Children corpus and the fourth-most in the BNC2014 children subcorpus, accounting for 127 instances of *yeah* (1.2 per 100 clauses) and 185 (0.8 per 100 clauses) respectively. In the Ar-Children corpus, confirmation marker *yeah* in turn-final position often carries a falling tone. An example of confirmation marker *yeah* in final position, with this intonation, is provided in (60).

(60) 085301: my favourite movie is Spider Man no way home/

085302: yeah. I watched it. it was so cool.

085301: it was so good. ***yeah.*** (0853/ Ar-Children)

Alternatively, *yeah* can express confirmation of the speaker's understanding of another speaker's comments. In the Ar-Children corpus, speakers use *yeah* in this way in 29 instances. For instance, in (61) Speaker 110101 employs *yeah* with *I know* to confirm her understanding to the other speaker.

(61) 110102: […] or-or there's the same mall / there's a place where snow, but / it's not real, it's fake,

110101: ***yeah*** I know, my mum told me about it. (1101/ Ar-Children)

. There are 19 instances of *yeah* in the Ar-Children corpus in which speakers confirm their recollection of past information after DM *oh*, as in example (62), or with expressions such as *I remember* as in (63).

(62) 094601: I also love mm Snow white,

094602: ***oh/ yeah*** Snow white. Snow white looks like Alayna, (0946/ Ar-Children)

(63) 958602: okay and, how many friends do you have.

958601: uh I have two, I have uh [ANON-name], and I have [ANON-name],

958602: oh okay,

958601: and [ANON-name] , yeah , I had also another one but she moved on , [ANON-name] moved on to other school , [ANON-name] , if you remember her , but she was.

958602: ***oh yeah*** / I remember her. (0958/ Ar-Children).

6.4.4.3. *Yeah* as an agreement marker

This function is associated with the core semantics of *yeah*, as speakers express their agreement with another speaker's comments. In the Ar-Children corpus, *yeah* has this function in 88 instances (0.8 per 100 clauses) compared to 135 instances (1.0 per 100 clauses) in the BNC2104 children subcorpus. Speakers often repeat another speaker's previous turn, or part of it, when using agreement marker *yeah*. Alternatively, they may use *yeah* alongside agreement expressions such as *me too*, *also me*, *same* and *I agree*. The agreement marker *yeah* often appears with a raising tone. In (64), from the Ar-Children corpus, speaker 092901 repeats the last part of her friend's turn to show agreement, and then offers her own explanation.

(64) 092902: or like Cappy. that's my favourite tag. but/ some of them are like , pricey,

092901: ***yeah/*** some of them are actually really pricey, like, sometimes you ca n't get it / (0929/ Ar-Children)

Speakers sometimes express *partial* agreement with other speakers by using *yeah* alongside other DMs such as *but* and *well* – which can indicate a speaker's desire to deny a previous presupposition or raise an objection (Norroik, 2009:874). For instance, in (65), speaker 094602 uses *yeah* followed immediately by *but* to show that she does not totally agree with her friend's opinion on dogs. The partial agreement shows that speakers are attentive to potential threats to positive face; 094602, for example, uses *yeah* before *but* as a softening device prior to the face-threatening act of disagreeing with 094601.

(65) 094601: I love dogs but like, small ones but the big ones,

094602: yeah, ***yeah*** but/ I like big ones too because, I do n't know, I 'm just afraid of cats, but I kinda like dogs, (0946/ Ar-Children)

6.4.4.4. *Yeah* as a speech act marker

Yeah as a speech act marker is infrequent in both corpora (0.1 per 100 clauses). As a speech act marker, *yeah* precedes questions intended to clarify or confirm previous information. In (66), speaker 124502 ends her turn with *yeah* to inform speaker 124501 that she has nothing to add on the current topic. Speaker 124501 continues with further ideas on the same theme. She employs *yeah* to precede her request for clarification of her friend's last turn.

(66) 124502: playing with my toys, and like, yeah

124501: ***yeah,*** what does you play it/ in the phone, play it in real, what do you like. (1245/ Ar-Children)

Speakers sometimes employ *yeah* to aid them in achieving collaboration in their interaction, though confirmation requests. For instance, in (67) speaker 091801 requests from her friend an explanation of a famous game that is played in her friend's country. She uses *yeah* in several turns to encourage her friend to continue with her explanation (see section 6.4.4.6). In her last turn, she employs *yeah* to introduce her confirmatory question.

(67) 091801: I can't understand, I do n't understand,

091802: when there are three players/

091801: yeah,

091802: one catcher,

091801: yeah,

091802: two players,

091801: ***yeah***, is it like the game, where you sit down, and you can't be catched/

(0918/ Ar-Children)

6.4.4.5. *Yeah* as an expression of emotion/opinion

Yeah can mark a speaker's expression of feelings and emotions using verbs such as *love, hate, like* and *(don't) care*, or adjectives such as *excited, happy, sad, fun(ny), amazing* or *nice*. There are 19 instances of *yeah* in the Ar-Children corpus that precede a speaker's expression of their feelings, as exemplified in (68).

(68) 084801: oh my gosh, yeah. and also we're planning your birthday,

084802: ***yeah***. I'm really excited as well. (0848/ Ar-Children)

In the BNC2014 children subcorpus, speakers similarly use *yeah* with this function. In addition, there are 13 instances of *yeah* in the BNC2014 children subcorpus in which speakers employ *yeah* to mark the expression of an opinion, using verbs such as *think, believe*

and *guess*. In (69), for instance, *yeah* marks speaker S0416's expression of an opinion about how to raise their younger brother.

(69) S0416: >> ***yeah*** I think you have to () teach --ANONnameM we have to teach --ANONnameM humble things (SV5A/ BNC2014 children)

6.4.4.6. *Yeah* as a backchannel marker

The backchannel function of *yeah* is its most frequent function in both corpora, accounting for 159 instances in the Ar-Children corpus (1.5 per 100 clauses, 88.7% of DM *yeah*) and 644 in the BNC2014 children subcorpus (2.8 per 100 clauses, 86.4% of DM *yeah*). In this function, *yeah* appears as a standalone marker to express a participant's intention to remain a hearer and not to take over the conversational floor (see section 6.2.4.4 and 6.3.4.5). It thus indicates that its producer is still listening, though it does not necessarily entail their agreement (Fuller, 2003a: 38). In the Ar-Children corpus, standalone *yeah* with this function may carry rising tone, as in (70).

(70) 103701: you know. mm four years ago uh خالي / bring us a rabbit/ from his friend's mm farm.
103702: ***yeah/***
103701: we played it for-for about- played with it, for about a day and then it was gone, we we were told that, uh uh it was dead (1037/ Ar-Children)

6.5. Summary

In this chapter, I focused on *okay*, *oh* and *yeah*, looking at findings about these DMs in prior studies before proceeding to quantitative and qualitative analysis. These DMs are among the most frequent in the two corpora, and function predominantly as interpersonal

markers in both. My analysis of these three DMs forms part of my answer to the third RQ of this study.

In section 6.2, I looked at the functions of *okay* in the two corpora. As a DM, *okay* is more frequent in the Ar-Children corpus than in the BNC2014 children subcorpus. The analysis shows that *okay* is predominantly an interpersonal marker in both corpora. Speakers employ *okay* to manage their turn, to confirm prior turns, to request further clarification or confirmation, and to request continuation. As a textual marker, *okay* has transitional and filler functions in both corpora. *Okay* as an initiator appears exclusively in the Ar-Children corpus, albeit infrequently. Nevertheless, initiator *okay* shows speakers' management of the conversational task and turn allocation.

Section 6.3 moved on to the DM *oh*. The overall frequencies show that *oh* is often an interpersonal marker in both corpora. In the interpersonal domain, *oh* is frequently used as a responsive marker, or to show understanding of previous turns; this is true of both corpora. It also marks confirmation and clarification requests and has backchannel functions. As a textual marker, the most frequent functions of *oh* are the initiator and transitional functions. My investigation of the transitional function confirms earlier findings that *oh* may be used in this way as a change-of-state marker.

In section 6.4, I discussed the functions of *yeah*, which is predominantly an interpersonal marker in both corpora (as earlier studies also found).

Overall, then, this chapter has shed light on the use of *okay*, *oh* and *yeah* as DMs in the English speech of Arabic-speaking children and native English-speaking children. The next chapter will proceed to discuss the results obtained in this and the prior two chapters, including certain issues noted but postponed in the analysis above.

Chapter 7: Conclusion

7.1. Overview

In this last chapter, I summarise the findings of my thesis, offer some discussions, and conclude the thesis. In section 7.2, I present the answers to my RQs (see section 1.3) based on the quantitative and qualitative corpus analysis regarding the most frequent DMs in the speech of Arab children learning English as L2 and English-speaking children. Then, in section 7.3, I discuss the implications of my findings for the study of DMs generally and for the exploration of pragmatic competence among learners of English as an additional language. In this section, I will also suggest potential explanations for the over or underrepresentation of DMs in the speech of L2 children. I will also discuss possible explanations for why Arabic-speaking children appear to rely on certain English DM functions, and how this relates to their pragmatic competence. In section 7.4, I will present some of the limitations of this study. In section 7.5, I will offer suggestions for future research based on the findings of the present study. In the last section 7.6, I will present the contribution of my thesis to the field.

7.2. Summary of findings

This thesis sought to examine the frequencies and functions of the most frequent DMs in the speech of Arabic-speaking school-age children learning English as an L2 and to compare these speakers' use of DMs with that of L1 children. In my review of previous research (see section 2.3.5), I found that researchers into DMs have investigated their use in the speech of language learners and compared them with that of native speakers (e.g. Romero-Trillo, 2002; Müller, 2005; Fung & Carter, 2007; Gilquin, 2008; Buysse, 2010, 2012, 2015; 2017; Buysse & Blanchard, 2022). However, my review demonstrated a lack of thorough investigation into children's acquisition of DMs in L2 contexts. My study investigated L2 children's acquisition of DMs as an indicator of their pragmatic competence.

In addition, it examined the role of context and length of preschool study on the use of English DMs by Arabic-speaking children. To fulfil these objectives, I addressed three RQs (see section 1.3). My findings regarding these questions were presented in chapters 4, 5, and 6. In this section, I will summarise these findings.

RQ1: *Are there any differences between the frequencies of the most frequent DMs in the ESL speech of Arabic-speaking upper-grade children (fourth, fifth and sixth grades) in Saudi international primary schools, compared to age-matched native English-speaking children?*

To answer this first RQ, I examined the wordlists in my two constructed corpora: the Ar-Children corpus and the BNC2014 children subcorpus (see sections 3.2 and 3.3). These wordlists showed that six forms, *like*, *yeah*, *okay*, *so*, *oh*, and *well*, were the most frequent forms of DM (see section 4.2). I supplemented the resulting list with two-word DMs which are frequently studied in L2 studies (see section 3.5.1). Additionally, I manually disambiguated the use of these forms as DMs from their non-DM use. The results showed that the eight forms (*like*, *yeah*, *okay*, *so*, *oh*, *well*, *you know*, and *I mean*) function mainly as DMs in the two corpora (see section 4.2). The differences between the two corpora in use of these eight DMs were statistically significant. By comparing the DM relative frequencies between the two corpora, I found that *like* and *okay* were more frequent in the Ar-Children corpus than in the BNC2014 children subcorpus. By contrast, *yeah* and *oh* were more frequent in the BNC2014 children subcorpus. *So* and *you know* appeared to be fairly similar in the two corpora. Although it was infrequent in both corpora, *I mean* occurred more in frequency in the BNC2014 children subcorpus. Since the comparison focuses on the overall group-level patterns of DM use, we cannot conclude whether individuals use certain DMs more frequently than others. I will discuss this limitation further in section 7.4.

The answer to RQ1 provides insights into the differences in DM frequency between L1 and L2 children. Indeed, the results on DM frequency as repeated linguistic patterns across the two corpora offer evidence of form-function pairings from a usage-based theoretical perspective. According to Bybee (2006: 7), investigating frequencies of linguistic patterns in child language can provide evidence of how children construct patterns that they have acquired based on their experience with adults including caregivers. In fact, the findings from my L1 dataset may reflect how the identified patterns of DM use have become entrenched through exposure and social interaction (see section 2.2.3). Similarly, the infrequent use of some DMs in the L2 dataset illustrates how L2 children are still acquiring these DMs. On the other hand, their frequent use of some DMs shows that those specific DMs have started to be routinized in their speech (see Tomasello, 2003, Bybee, 2006, 2010). I will expand on the possible reasons for the high and low frequencies of different DMs in each dataset (e.g. age, L1 transfer) in section 7.3.1.

RQ2: *How does the context of formal exposure to English before joining primary school affect the frequency of DMs in the ESL speech of these Arabic-speaking children?*

This second research question consisted of two sub-questions. The first sub-question was as follows: *is there any difference in the frequencies of DMs between participants who have studied ESL in preschool in Arabic-speaking versus English-speaking countries?*

To answer the first sub-question, I compared the use of DMs between participants who attended preschool in Arabic-speaking countries and those who attended preschool in English-speaking countries, by creating four different subcorpora (see section 3.4). My analysis showed no evidence of a positive effect on the frequency of DMs from attending preschools in English-speaking countries (section 4.3). Arabic-speaking children who attended preschool in Arabic-speaking countries produced DMs with frequencies similar to those of their peers who attended preschool in English-speaking countries, regardless of the

duration of preschool attendance. However, participants who attended preschools in Arabic-speaking countries used less varied DM types than those who attended preschools in English-speaking countries.

The second sub-question was as follows: *Is there any effect of the number of years of attendance at preschool in English-speaking countries on the frequencies of DMs?*

To answer the second sub-question, I compared participants' use of DMs according to the time spent in preschools in English-speaking and Arabic-speaking countries. I also tried to control for the age of participants as one of the factors that might also affect the use of DMs. My analysis showed different levels of DM use among the various age groups. For instance, in the 9-10-year-old group, I found that *yeah*, *oh,* and *okay* are less frequent in the AP2 group (two-years experience in English-speaking countries), but on the other hand, *like*, *so*, *you know,* and *I mean* are used more frequently by the AP1 group (one-year experience in English-speaking countries). In the 11-12-year-old group, on the other hand, the relative frequencies of *like*, *okay,* and *yeah* were high in the AP1 group and lower in the AP2 group. The results for the 11-12-year-old group also revealed that the more time spent in English-speaking preschools, the more Arab children used the less frequent DMs *well,* and *I mean.* I analysed the difference in frequency profiles among the three groups based on their preschool experience and found significant differences across different age cohorts. The significance tests' results suggests that preschool experience in different language contexts may play a role in shaping the frequency patterns of DMs as Arabic-speaking children develop their language skills. However, the contradictory results between the two age cohorts make it difficult to reach firm conclusions. Because of the choice of statistical test, my findings of the different groups must be interpreted as reflecting aggregate data rather than any property of given individuals' use of DMs. I reflect further on this limitation in section 7.4.

The answer to the second RQ may enhance our understanding of how the use of DMs are affected by both internal and external factors. Examining the effect of these factors supports the usage-based approach to language acquisition. Indeed, the frequency of linguistic items, and the entrenchment of language patterns, are often shaped by such factors (Unsworth, 2016a, see also Blom et al., 2012). However, the contradictory findings that emerged for the two age groups makes it difficult to establish a clear link between external factors and the frequency of DM use in my data. I will discuss this further in section 7.3.3.

RQ3: *What are the functions of the most frequent DMs in the speech of these Arabic-speaking school-age children? Are there any differences in the functions with which DMs are used, in comparison to native English-speaking children?*

To answer the third RQ, I manually coded the eight DMs according to the two domains of functions (i.e. textual and interpersonal) in the proposed framework of DM functions (section 3.5.2). In the Ar-Children corpus, the eight DMs more often had textual functions (61.1%) than interpersonal functions (38.9%). The percentages of the two domains were reversed in the BNC2014 children subcorpus (47.1% for textual and 52.9 % for interpersonal functions). The differences between the two corpora regarding the two domains of functions were statistically significant for the two DMs *okay* and *so*. In contrast, no differences were found between the two domains of function across the two corpora for the other six DMs: *like*, *yeah*, *well*, *oh*, *you know* and *I mean*.

Looking at individual DMs, I observed differences in functions between the two corpora. For instance, *well* was mainly an interpersonal DM in the Ar-Children corpus but mostly a textual DM in the BNC2014 children's subcorpus (section 4.5). This difference was statistically significant. In the Ar-Children corpus, *well* was frequently employed as a filler, transitional, responsive, or agreement/disagreement marker. In comparison, *well* in the

BNC2014 children subcorpus often functioned as a filler, elaborator, or responsive marker (section 4.5.2). The rare use of *well* in the Ar-Children corpus showed that the Arabic-speaking children have not (yet) acquired the full range of *well*'s functions as a DM (see further section 7.3).

DMs, *you know,* and *I mean* were more frequently used to mark textual functions than to mark interpersonal functions (section 4.6 and 4.7). The differences were statistically significant, however, no equivalent significant difference was found for *I mean*. Both *you know* and *I mean* carry various textual functions in the two corpora, including filler and elaborator functions. Initiator *you know*, which was also observed, was more frequent in the Ar-Children corpus than in the BNC2014 children subcorpus. While *you know* carried various interpersonal functions, including turn management, shared knowledge, and speech act functions, *I mean* was used mainly to flag turn management interpersonal functions.

Like is mainly a textual DM in the two corpora, with three functions: initiator, elaborator, and filler (section 5.2.2). These functions were three times more frequent in the Ar-Children corpus than in the BNC2014 children subcorpus. The differences between the functions of *like* in the two corpora were statistically significant. I did code some examples of *like* as having interpersonal functions (section 5.2.4). However, these functions were not frequent enough to merit separate categories. This result aligns with prior studies of *like* carrying interpersonal function (see Miller & Weinert, 1995, Diskin, 2017).

Certain subfunctions of elaborator *like* occurred in the Ar-Children corpus only, such as use of elaborator *like* alongside code-switching (see section. 5.2.3.2). This use of elaborator *like* was not frequent enough to formalise as a separate category. Nevertheless, these observations add to our understanding of how speakers can utilise code-switching to maintain discourse continuity.

So carried more textual than interpersonal functions in both corpora (see section 5.3.2). As a textual marker, *so* marked various functions, including initiator, filler, transitional, and elaborator functions in the two corpora. The differences between the functions of *so* in the two corpora were statistically significant. Participants in the Ar-Children corpus used initiator *so* more frequently than other textual functions, whereas participants in the BNC2014 children subcorpus more often relied on transitional *so*. In both corpora, speakers use DM *so* to signal three different interpersonal functions: speech act, conclusion, and turn management functions.

The frequencies of *okay's* functions vary between the two corpora (see section 6.2). The number of occurrences of these functions was significantly different in the two corpora. *Okay* carried textual functions, including initiator, filler, and transitional in both corpora (section 6.2.2). Transitional *okay* was more prevalent in the Ar-Children corpus than the other textual functions. In the interpersonal domain, *okay* marked functions such as turn management, confirmation, speech act, and backchannel functions in both corpora.

Oh was mainly an interpersonal marker in both corpora, having speech act, understanding, emotion, responsiveness, and backchannel functions. In addition, *oh*, in both corpora marked the four main functions in the textual domain, i.e. initiator, elaborator, filler and transitional functions. The different textual and interpersonal functions were more frequent in the BNC2014 children subcorpus than in the Ar-Children corpus (section 6.3.2). The differences between the frequencies of *oh's* functions in the two corpora were statistically significant.

Over 80% instances of *yeah* in the two corpora carried interpersonal functions (section 6.4.2). In both corpora, speakers employed *yeah* for various interpersonal functions such as turn management, confirmation, agreement, speech act, and backchannel functions. As a textual DM, *yeah* had elaborator, filler, and transitional functions in both corpora. The

frequencies of these textual functions were significantly higher in the BNC2014 children subcorpus than in the Ar-Children corpus. At this point I must reiterate the critical point that these findings reflect group-level use of DMs in the two datasets, rather than trends across individual differences; see further section 7.4.

Considering the results for RQ3 in light of the theoretical assumptions of the usage-based framework suggests that both the L1 and L2 children represented in my data have acquired many of the discourse and pragmatic functions of DMs. According to this framework, children's acquisition of discourse elements and constructions often depends on how complex and frequent these elements are. Tomasello (2003: 265) observes that in child L1 acquisition, certain patterns can be observed in the learning of discourse elements such as connectives. Specifically, children tend to acquire discourse elements tied to their simple uses (e.g. conjunctions such as *and*, *or*) earlier, not tied to uses that are more complex and less frequent (e.g. connectives expressing psychological states of cause or reasoning such as *because*, *so*). The tendency of acquisition of sets of linguistic items to occur in particular orders across different children is a useful interpretive frame for the findings from my data. Although the L1 and L2 children in my datasets demonstrate the use of many DM functions, logically we might expect them to be still be in the process of acquiring the more complex DM functions (since those would be at the late end of whatever consistent sequence of acquisition exists for DMs and DM functions). On the other hand, children's acquisition of DM functions may exhibit individual differences due to internal factors such as cognitive maturation and age, as well as external factors such as frequency of exposure (see section 2.2.1). I will further discuss the factors that may explain the use, or lack thereof, of certain DM functions across the two datasets in section 7.3.2.

7.3. Discussion

7.3.1. Over/underrepresentation of DMs in the speech of L1 and L2 children

As noted earlier, *like* was the most frequent DM in both corpora. This result aligns with previous studies that find *like* to be overrepresented in the speech of young L1 people (under the age of 13) in modern-day English (see Andersen, 2001; Daily-O'Cain, 2000; Levey, 2006; Odato, 2013; Beeching, 2016; Schweinberger, 2014, 2023 among others). However, speakers in the Ar-Children corpus employ *like* more often compared to the BNC2014 children subcorpus. Its high frequency in the Ar-Children corpus suggests that L2 children can use *like* as a DM, similarly to their L1 counterparts. Indeed, participants in my data employ *like* to mark various functions almost equally (initiators 34.9%, elaborator 29.9, fillers 33.3%). These results show that Arabic-speaking children are aware of an English discourse pattern (namely, frequent use of *like* as DM), an awareness that could be a sign of pragmatic competence in their L2. Looking at the use of *like* by adult L2 speakers, Lorenz (2022) observes that DM *like* becomes more prevalent in the speech of advanced L2 learners compared to less experienced learners. According to Lorenz, the reason for this prevalence is the use of media, which makes it easier for learners to reach the global world and get in touch with L1 speakers (Lorenz, 2022). Although measuring the effect of media was not one of the objectives of my study, I noticed that participants, in most of the conversations, were talking about chat video games, social media applications, or English TV series. This informal observation may add to our understanding of why participants in the Ar-children corpus employ the DM *like*, in particular, more than other DMs (see section 4.2). Another interpretation for the overreliance on *like* over other DMs in the Ar-Children corpus could be the effect of peers' language. According to Schweinberger (2023), L1 children's use of *like* increases when they join schools due to their interactions with peers. In my data, participants

were all upper primary school students, at an age when children become peer-centred (see section 3.2.1). In this context, I propose that it might also be the effect of peers' language that leads to the high frequency of *like* as a DM.

The overreliance on *like* in the Ar-Children corpus may also be a negative sign of their L2 acquistion. Buysse (2015) asserts that L2 learners often rely on DMs that they are familiar with and tend to use "less complex pragmatic devices". This interpretation might explain why Arabic-speaking children use *like* over other DMs. In addition, it could be a sign for their limited proficiency in L2. Huang et al (2023) find that low-level learners use *like* more frequently than advanced learners, contradicting the findings of Lorenz (2022), discussed earlier. However, this discrepancy sheds lights on the potential effect of other factors (e.g. L1 transfer, L2 proficiency) that should be considered. Ultimately, both interpretations of L2 children's overreliance on *like* over other DMs remain inconclusive and warrant further systematic investigation.

Looking at the frequencies of other DMs in the Ar-Children corpus, I observed that Arabic-speaking children employed *okay* more often than speakers in the BNC2014 children subcorpus. One explanation may be that *okay* is frequently used as a loanword in spoken Arabic (Aljutaily, 2021). Moreover, English DM *okay* is translation-equivalent to the multifunctional Arabic DM *tayeb* 'good, okay, fine' (Al-Harahsheh & Kanakri, 2013; Aljutaily, 2021). The loanword *okay* is often used in spoken Arabic instead of *tayeb* a DM. Arya (2022) observes that Thai-speaking learners of English's overreliance on *okay* results from it being a loan DM in Thai; the explanation I propose here is that the same applies, in parallel, to Arabic-speaking learners.

Other DMs in the Ar-Children corpus were less frequent than in the L1 data, e.g., *well* (0.8 instances per 100 clauses). The low frequency of *well* in the Ar-Children corpus compared to the BNC2014 children subcorpus suggests that these children have not yet

acquired its function as a DM. Indeed, *well* is primarily connected with interpersonal functions, including turn management and responsive markers (see section 4.5.1). Although speakers in the Ar-Children corpus do use these functions with other DMs, these two functions in particular are infrequent with *well* in the Ar-Children corpus. The Arabic-speaking children employ the turn management function with other DMs, including *yeah* (1.5 per 100 clauses) and *okay* (0.9 per 100 clauses); likewise, they use the responsive function more often with DM *oh* (0.6 per 100 clauses).

The underrepresentation of *well* in the Ar-Children corpus contrasts with some previous studies, which found *well* to be more frequent in the speech of L2 than the speech of L1 speakers (Gilquin, 2008; Aijmer, 2011; Buysse, 2010; Huang, 2011). These findings could be attributed to the L1 effect in this case, including whether or not *well* has a translation equivalent in L1. Some of these studies report that *well* is used less frequently by Chinese learners of English than by other groups of learners, including native speakers of French, Spanish, and Dutch (Buysse, 2015), Swedish (Huang, 2019) or native speakers of English (Liao, 2009; Bu, 2013). Although none of these studies investigates the role of L1 interference on the acquisition of DMs, their results, in combination with those of the present study, should direct our attention to the potential role of such L1 effects. Notably, these earlier results come from studies of adult learners' use of *well*. However, this feature is known to differ between adults and children. Beeching (2016: 61) observes that young English speakers (0–14-year-olds) use *well* less frequently than other groups of speakers. Likewise, Romero-Trillo (2002: 781) reports that L1 and L2 children (first graders) use *well* infrequently compared to adult speakers. These results suggest that *well* is not fully acquired until adolescence. In my study, *well* is one of the 50 most frequent words in the BNC2014 children subcorpus. Its high frequency in L1 data shows that L1 speakers can use *well* as a DM from a young age. The low frequency of *well* in the Ar-Children corpus, therefore, could

be a sign that Arabic-speaking children have not yet acquired the functions of *well* as a DM in English. This finding warrants further investigation.

The low frequency of the two DMs *you know* and *I mean* is not confined to the Ar-Children corpus. In both my corpora, speakers employ these two-word DMs only rarely relative to other DMs. The low frequency of these two markers in L2 children's speech tallies with previous studies that report a scarcity of *you know,* and *I mean* in the speech of L2 adults (Müller, 2005; Fung &Carter, 2007; Hellermann & Vergun, 2007; Mei, 2012; Buysse, 2011, 2017) and children (Romero-Trillo, 2002). The infrequent use of *you know,* and *I mean* by L1 speakers in the BNC2014 children subcorpus, although contrasting with prior studies that found these markers to be prevalent in the speech of L1 adults and children (Schiffrin, 1987; Erman, 2001; Beeching, 2016; Buysse, 2017), may provide insights into L1 children's acquisition of these DMs. For instance, previous studies have established that these two DMs are not acquired by adolescence and often appear in the speech of middle and older people (Macaulay, 2002; Beeching, 2016). According to Andersen's proposition (2001:304), young people prefer to stick to simple form of DMs (i.e. one word DMs) due to cognitive constraints. It is relevant that school-age children are still in the process of acquiring complex syntactic structure and discourse skills as they start to learn other skills, i.e. literacy (Nippold, 2004: 6). Acquiring these skills often needs time and social engagement with family members and peers (Blum-Kulka & Snow, 2004: 4). *You know* and *I mean* often carry interpersonal functions (see section 4.6.1 and 4.7.2), which adds to the complexity of their acquisition (Östman, 1981:65). From the low frequencies of *well*, *you know,* and *I mean,* we can conclude that the school-age children in the present study are still developing their language repertoire, including use of DMs.

7.3.2. Functions of DMs in the speech of L1 and L2 children

Before looking for specific differences between the two corpora, I will discuss the overall picture regarding the differences between the two main domains: textual and interpersonal. For instance, Arabic-speaking children as L2 learners use the eight DMs mainly to perform textual functions (61.1%). By comparison, the percentage difference between domains is less marked in the speech of L1 children (47.1% for textual and 52.9 for interpersonal). There was no statistically significant difference between the functions of the six DMs *like, yeah, oh, well, you know* and *I mean*. The only two DMs that showed significant differences were *so* and *okay*. The L2 children's excessive use of textual functions of *so* and *okay* may result from still being in the learning phase (see section 7.3.1). Polat (2011) argues that an L2 speaker experiences a phase of excessive use of certain functions (filler) that then declines as they become more advanced learners. Looking at the results for *you know* as one example (see section 4.6), we find that it is used as a turn management marker 9.3% of the time, whereas in the L1 children's speech, it is used that way 22.8% of the time. This difference is statistically significant between the two corpora. In addition, L2 speakers stick more frequently to other DMs when they attempt to take over the conversational floor (*yeah, okay,* and *I mean*). Together, this suggests that the Arabic-speaking children can use this function but have not yet mastered it with the DM *you know*. Many previous studies observe that L2 adult learners show a developmental path in acquiring the functions of DMs. Those functions related to interaction with others, i.e. interpersonal functions, are usually acquired later than textual functions (Romero-Trillo, 2002; Liu, 2016; Ament & Parés, 2018). For example, based on observation of L1 children, Östman (1981) argues that there is a hierarchy of complexity in acquiring DMs. DM functions linked to the relationship between speakers and hearers are more complex to develop than those that signal coherence relations (Östman, 1981: 65).

In looking at the main functions of individual DMs, I found no differences between L1 and L2 children. Instead, only slight differences in their use of subfunctions were noted. For instance, I observed that Arabic-speaking children used the initiator function of *like* to highlight new information more than the other functions of *like* (34.9%). Andersen (2001: 247) argues that young speakers like to display uncertainty about new information and its appropriateness. This explanation would be congruent with Müller's (2004:1175) finding that L2 learners prefer explicitly displaying their "lack of confidence in their linguistic abilities" when communicating in their L2. Also, according to Schweinberger (2023: 20), young L1 children often use *like* to show uncertainty in their interactions. It is possible that L2 learners are still in the process of learning these various functions of *like* including its use to express uncertainty.

Okay is another DM for which the frequencies of functions differ between the two corpora. A high proportion of *okay* in the BNC2014 children subcorpus serves backchannel and turn management functions (29.2% and 26.5%, respectively). This aligns with prior research on *okay* as an acknowledgement marker and marker of turn-taking (Beach, 1993, 2020). By contrast, in the Ar-Children corpus, *okay* is used with roughly equal percentage as a filler (17.3%), turn management marker (17.6%), and confirmation marker (17.7 %). This result shows that children in my data demonstrate the use of *okay* as a turn-management marker. Baines and Howe (2010) observe that the nine-year-old children in their study had developed topic management skills and became aware of collaborative discussion strategies. They concluded that as children age, they become more pragmatically competent in using different conversational management devices, including DMs. However, the high frequency of *okay* as a filler in the Ar-Children corpus shows that the L2 speakers are using this DM as a "dummy marker" (a term suggested by Buysse 2015 for overreliance on a DM to perform

filler functions). This overreliance on *okay* as filler in the Ar-Children corpus could result from its use as a loanword in Arabic, as previously discussed.

Observing the use of filler functions in both datasets shows that L2 children employ this function differently than L1 speakers. While using DMs to flag self-repair is quite similar in frequency across the two corpora (1.4 per 100 clauses in the Ar-Children and 1.1 per 100 clauses in the BNC2014 children subcorpus), the use of planning filler functions differs between the two corpora. L2 children in the Ar-Children corpus use DMs with planning filler function to signal their hesitation, their thinking process or a search for words more frequently than do the L1 children in the BNC2014 children subcorpus (4.2 per 100 clauses in the Ar-Children corpus and 2.2 per 100 clauses in the BNC2014 children subcorpus). The appearance of DMs alongside filled and unfilled pauses to show filler functions in both corpora adds to our understanding of how the acquisition of DMs may be shaped by usage frequency and interactional experience. According to Tomasello (2003), language acquisition is based primarily on children's ability to extract linguistic patterns from repeated instances of usage in social context. Thus, the co-occurrence of DMs with filled pauses reflects form-function mappings that support the learning of pragmatic functions such as planning and self-repair. Ellis (2002) further suggests that L2 learners often acquire language constructions through exposure to patterns that serve their communicative purposes. According to my results, both L1 and L2 use DMs to serve the functions of planning, hesitation and search for words more frequently than they use these markers to flag self-repair. L1 and L2 children are able to use DMs to flag their need to plan their upcoming discourse which, therefore, serves their communicative purpose. However, the infrequent use of filler DMs to flag self-repair shows that both the L1 and L2 children in my data are still learning how to use the various filler functions of DMs.

7.3.3. The effect of attending preschool in English-speaking countries

In my study, I tried to add to the existing knowledge about the effect of attending preschool in English-speaking countries on the acquisition of DMs, following prior studies that report a positive impact of such preschool attendance on various linguistic aspects, including morphology and syntax (LIanes & Muñoz, 2012; Paradis, 2017). Analysis of the DM frequencies according to the context and duration of preschool attendance in an English-speaking country showed no difference in the overall use of the eight DMs (section 4.3). However, variation was observed in the frequencies of some DMs between the groups of speakers. Contradictory results from the two age groups make it difficult to reach firm conclusions regarding the effect of the context and duration of formal exposure on the use of DMs by Arabic-speaking children. However, the variation of DM use among the three groups may shed light on the possible effect of the context of learning DMs on children's use of theses markers. Such factors warrant further investigations in future studies.

7.4. Limitations

This study tried to shed light on the differences between L2 Arabic-speaking children and L1 children's use of English DMs. However, due to constrained time and space, some limitations were unavoidable. For instance, before collecting my data, I aimed to investigate the use of DMs in L2 children generally (see section 2.2.1). However, I had to make certain decisions regarding what participants to include in the present study. This means controlling for certain external factors. The first decision was to include only female children due to the practicalities of collecting data in the context of English-medium international schools in Saudi Arabia (see section 3.2.1.2). Controlling such factors was advantageous in one way as it ensured the homogeneity of participants. However, it prevented me from looking into salient differences in use of DMs between different speaker groups, in this case the two

genders. Researchers to date have observed small or insignificant differences between males and females in use of DMs (Escalera, 2009; Beeching, 2016; Huang, 2019; Lorenz, 2022). For instance, Beeching (2016) observes only slight differences between male and female speakers in the BNC in use of DM functions, but not in DM frequencies. In the present study, I did not consider gender differences as a factor in DM acquisition because my data focused on female children only as stated earlier.

The other decision I made before data collection was to limit the age of participants and include only upper primary students (see section 3.2.1). Collecting data from different age ranges would help examine how L2 children acquire DMs across the years of their development. This would entail a considerable amount of data collection which may require more time, materials and management including recruiting more participants, intensive data collection and complex data analysis. I decided to focus on only one age range due to time constraints. This decision enabled me to get more focused data from one age range rather than collecting small datasets across various age groups.

The second limitation is also related to the data collection process. My aim in compiling the corpus of L2 child speech was to collect spontaneous and natural conversations. However, some aspects of the collection methods may have limited that spontaneity and naturalness. Although I did not participate in the discussions, the participants knew I was present in the room. This may have had an impact on how the interactions developed. For instance, participants may have been more conscious of their choice of topics than they would have been without any observer present. Some participants (although willing to participate) were reluctant to speak, and needed encouragement. Their hesitation could have led to unnatural data. I tried not to intervene in their discussions to preserve the spontaneity of conversations (see section 3.2.2.1). Most participants carried out their

conversations without any need for assistance and without paying any apparent attention to my presence as a researcher.

The third limitation relates to the dataset that I used as a point of comparison for the Ar-Children corpus. As discussed in section 3.3, the spoken BNC2014 was the optimal choice for my comparable data based on issues of availability and compatibility, even though there were still quite serious points of mismatch. But these points of mismatch may well have limit the extent to which the results of my study can be generalised. The two points of mismatch most notable here are the context of the child speech collected, and the lack of prosodic information in the Spoken BNC2014. First, unlike the Ar-Children corpus, which consists of child peer conversations, the child utterances which I drew from the Spoken BNC2014 come from conversations between adults and children. Previous studies have found that differences between types of interactions do not necessarily lead to differences in usage of DMs (see section 2.3.5). However, the possibility remains that such differences may in fact exist. I have at all stages explicitly acknowledged that the differences between the asymmetrical interactions from which the BNC2014 children subcorpus was drawn and the symmetrical interactions in the Ar-Children corpus may affect the frequencies of DMs that are to be found in each dataset. So the use of the Spoken BNC2014, although optimal in the sense that all other possible comparable datasets presented even greater problems of one sort or another, creates a substantial limitation in the extent to which comparative patterns in my data can be relied upon.

The other point of mismatch is the lack of prosodic information in the spoken BNC2014, which complicated some of the distinctions I needed to make among DM functions. This was not the case for my own corpus, which was prosodically annotated. Although the L1 data was not annotated, I managed to code the functions of DMs relying on extended prior context and on adjacent cues, such as filled and unfilled pauses and agreement

from other speakers (see sections 4.6.3, 5.3.4, 6.3.4, and 6.4.4). In case of a mismatch, I opted to code the occurrence of the DM as unclear. Despite this limitation, the Spoken BNC2014 corpus remained the optimal choice for the present study. This is because the Spoken BNC2014 corpus was compiled using spontaneous conversations between family members and friends. In addition, children's participation in the corpus provides a substantial number of tokens, facilitating meaningful comparisons with the L2 corpus (see section 3.3).

The fourth limitation relates to my use of the theoretical framework of usage-based. The main objective of my study was to investigate the frequencies and functions of DMs in the speech of L2 children compared to L1 children through the use of corpus methods. The usage-based framework aligns perfectly with this objective. But my focus on investigating the frequency of DM mainly through corpus-based approaches meant that I omitted detailed analysis of the effect on DMs of other factors that are, from a usage-based theoretical view, highly relevant to children's acquisition of any kind of linguistic item. One of these factors is the effect of priming, that is, increases in the likelihood that a speaker will produce an item or pattern resulting from prior production or perception of that item or pattern (see section 2.2.3). However, the relevant literature indicates that investigating the effect of priming can best be achieved via carefully designed experimental tasks. To undertake this kind of detailed experimental research *as well as* constructing a novel corpus of L2 children's speech would have been far beyond the scope of a single PhD project; given my goal of studying frequency, the choice to invest effort in corpus construction rather than experimental investigations was clear.

The fifth limitation relates to choices regarding the statistical analysis of DM frequencies. I chose to employ the Log-Likelihood test to compare the use of DMs across the two corpora – this test being particularly suitable to looking at the differences between L1 and L2 child speakers as collective groups rather than looking for individual differences. This

approach has its limitations. First, Log-Likelihood treats all occurrences of DMs as independent and ignores any clustering of some DM or DMs in the language produced by one or a few specific speakers. Indeed, a potential reason for any high frequency observed for a specific DM is its being extremely frequent in the usage of a single speaker's speech or in a single text (Gablasova et al., 2017: 148). Another limitation of Log-Likelihood is that it does not account for factors distinguishing different factors such as age, proficiency or amount of input. Additionally, examining individual differences through other statistical tests in the present study would be conceptually problematic due to one of the other limitations- namely, the differences in participant selection procedures and conversational context between L1 and L2 children (see sections 3.2 and 3.3). This type of test, however, was the most suitable choice in the present study because of its compatibility with my primary goal of group-level analysis focusing on discrepancies in DM frequencies between the two corpora.

Another limitation related to statistical analysis was choosing clause count as the basis of analysing relative frequencies in the two corpora (see section 3.5.3), although it is typical for researchers to employ word token counts for this purpose. This choice may arguably affect the generalizability and comparability of the findings. For instance, comparisons on the basis of clause counts may overlook phenomena such as DM clustering. Furthermore, other research on DMs has often relied on word/token counts; it may in consequence be difficult to compare the findings of the present study to earlier research. Yet despite all this, I believe that the use of clause counts was justified because it avoids overstating the significance of differences in the data by producing very large numbers based on token counts (see Stefanowisch & Gries, 2003). That is to say, clause count is a more conservative measure than the traditional word count. Therefore, differences found to be significant based on clause counts in the present study constitute a subset of the results that would have been found significant had I worked with word counts. This deduction is confirmed by the fact that I also

calculated the comparisons in question on the basis of word counts, and obtained similar results. Future studies could further test the robustness of the present findings by comparing clause count with word token count across larger datasets or different DM categories.

The last limitation relates to my assigning the label *filler* within the textual domain to the set of DM functions that flag processing or repair in the upcoming discourse. Traditionally, the term *filler* refers to words that occupy pauses or gaps in someone's speech. In my study, however, I defined the word *filler* differently (though not entirely originally, as I follow other scholars here; see section 3.5.2). By my definition, a *filler* is any word that marks a speaker's difficulty in processing upcoming discourse or flags an attempt to repair their own speech. Using this definition, I identified a set of functions in which DMs flag speakers' attempts to gain time to think, search for words or expressions, or repair themselves, and I grouped these all under the single label of *filler* functions. However, after the fact and in writing up my results, I found that assigning the label *filler* to these various functions had the negative effect of seeming to minimise the importance of certain pragmatic functions of DMs, such as self-repair, due to the implications of the traditional denotation of *filler* as extra or meaningless words. Although my framework's handling of the functions in question did serve the purposes of the present study, I suggest that this use of the term *filler* needs to be reconsidered in future studies, and possibly replaced by less potentially confusing terminology.

7.5. Directions for future research

In my study, I tried to look at the differences in use of DMs in the speech of Arab children concerning the time spent in preschool in English-speaking countries as an external factor. Future studies may examine the use of DMs concerning additional external and internal factors. Such factors include the amount of input during study abroad and the

intensity of formal education. For instance, children might be enrolled in preschool but not full-time. We might also consider the influence of exposure to social and broadcast media. In addition, future researchers may also be interested look at language-internal factors relevant to children's acquisition of DMs, such as the possible role of priming. These factors, alongside many others, should be considered in future studies investigating the language of school-age children and their pragmatic competence.

Before assigning the DM functions, I had to distinguish the DM use from non-DM use in the forms under study (see section 3.5). During the process of manual disambiguation, I observed some points unrelated to present concerns but worth highlighting for future research. For instance, tag questions in the Ar-Children corpus often did not appear in the canonical form of auxiliary verb plus optional negation plus pronoun co-referencing the main clause subject (Biber et al., 1999: 208). This canonical type of tag question was easily detected in the BNC2014 children subcorpus. The other form of tag questions, less strictly defined form than the first, appeared with DMs as question tags, for example, *okay* and *you know*. The second form, unlike the first, appeared in both corpora, as in example 1.

(1) 392901: before I came to the school. I was in [ANON-place], ***okay/***
392902: uhu (3929/ Ar-Children)

Another point I observed in the BNC2014 children subcorpus but not in the Ar-Children corpus was using *so* as a substitute for a predicate in the previous clause (Biber et al., 1999: 917), as in example (2).

(2) S0654: >>no () I ate peanuts M and M I ate some
S0653: >>right
S0655: >> oh, ***so*** did I (SZ6C/ BNC2014 children).

Although this example of *so* is not a DM, I encountered many of these examples while distinguishing *so* as DM from non-DM use- in the BNC2014 children subcorpus only. The absence of *so* as a substitute in the Ar-Children corpus suggests that researchers should consider L2 children acquistion of such grammatical structures.

As discussed earlier, some of my results suggest the possibility of an effect from language transfer, that is of translation-equivalent items in the L1 or lack thereof. For instance, the Arabic DMs **يعني** (y*aani* 'it means'), as in example (3), is equivalent to the English DM *I mean*.

(3) 090202: it's so fun, and I go skiing/ I had skiing practice. oh, what sport do you play/

090201: some words, somewhat

090202: sports, **يعني**

090201: sport /

The use of يعني was more frequent (0.4 per 100 clauses) than the use of *I mean* (0.2 per 100 clauses). Investigating Arabic DMs was not an aim in my study. However, Arabic-speaking children's use of these markers during their conversations in English might open a window of opportunity for further research on interlanguage pragmatic acquisition in Arab children and adults as well. Such studies could benefit from contrastive corpus analysis of comparable or parallel corpora.

The present study looked at the use of DMs in L1 and L2 children's speech according to a group-level use. This decision was made to serve certain purposes discussed earlier and to serve the objective of the study. However, it would be beneficial for future studies to consider the findings of the present studies (e.g. frequent DMs, frequent functions) in designing further studies that focus on how individual L2 children use these DMs. Moreover,

more future studies are needed to investigate how external and internal factors affect the use of DMs by individual L2 children.

Although the data in the Ar-Children corpus was transcribed, I was able to consult the audio when needed. But for the spoken BNC2014 children subcorpus, audio was unavailable. The lack of audio or prosodic annotation made distinguishing some DM functions difficult. Thus, there is a need either to annotate a pre-existing dataset such as the Spoken BNC2014 or to compile a corpus of informal L1 child speech that is prosodically annotated. Investigation of such a corpus could make a considerable contribution to L2 child acquisition research. In addition, future research would benefit from the availability of multimodal corpora in which researchers can both listen to the speakers and take into account gesture and gaze when analysing the functions of DMs.

7.6. Contribution

Although much research has employed the corpus method to examine the use of DMs, the present study contributes to the field of L2 acquisition as one of the first studies to focus on Arabic-speaking children's use of DMs in peer interactions. This is a group of speakers that has not been examined in other research. This study sheds light on the possible DM functions employed in the speech of Arabic-speaking school-age children and L1 children. These findings, hopefully, will help L2 teachers build materials to assist L2 children in learning the various functions of DMs. For instance, students can benefit from teachers' guidance in performing their conversational tasks by supplementing them with a list of corpus examples of DM use to enhance the spontaneity of their interactions. I hope this thesis may inspire future research in child SLA to address how L2 children from other cultural backgrounds use DMs in their interactions. More research could focus on L2 children's

pragmatic competence and how these children's learning development may differ from that of adults.

References

Adolphs, S. (2008). *Corpus and context: Investigating pragmatic functions in spoken discourse*. John Benjamins.

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. John Benjamins.

Aijmer, K. (Ed). (2009). *Corpora and language teaching*. John Benjamins.

Aijmer, K. (2011). *Well* I'm not sure I think… The use of *well* by non-native speakers. *International Journal of Corpus Linguistics*, 16 (1), 231–254. https://doi.org/10.1075/ijcl.16.2.04aij

Aijmer, K. (2013). *Understanding pragmatic markers in English*. Edinburgh University Press.

Aijmer, K. (2015). *Well* in an English-Swedish and English-French contrastive perspective. In K. Beeching & H. Woodfield (Eds), *Researching sociopragmatic variability: Perspectives from variational, interlanguage and contrastive pragmatics* (pp. 230–250). Palgrave Macmillan.

Aijmer, K. (2016). Pragmatic markers as constructions: The case of *anyway*. In G. Kaltenböck, E. Keizer, & A. Lohmann (Eds), *Outside the clause: Form and function of extra-clausal constituents* (pp. 29–57). John Benjamins.

Aijmer, K. (2018). Corpus pragmatics: From form to function. In A. H. Jucker, K. P. Schneider, & W. Bublitz (Eds), *Methods in pragmatics* (pp. 555–586). Mouton De Gruyter.

Aijmer, K. & Altenberg, B. (1991). Introduction. In K. Aijmer & B. Altenberg (Eds), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 1–7). Longman.

Aijmer, K. & Rühlemann, C. (Eds). (2015). *Corpus pragmatics: A handbook.* Cambridge University Press.

Aijmer, K. & Simon-Vandenbergen, A. (2006). Pragmatic markers in translation: A methodological proposal. In K. Fischer (Ed), *Approaches to discourse particles* (pp.1-30). Elsevier.

Al Makoshi, M. (2014). *Discourse markers and code-switching: Academic medical lectures in Saudi Arabia using English as the medium of instruction.* Unpublished PhD Thesis, University of Birmingham.

Albirini, A. (2018). The role of age of exposure to English in the development of Arabic as a heritage language in the United States. *Language Acquisition*, *25*(2), 178–196. https://doi.org/10.1080/10489223.2016.1234620

Alduais, A., Al-Qaderi, I & Alfadda, H. (2022). Pragmatic language development: Analysis of mapping knowledge domains on how Infants and children become pragmatically competent. *Children*, *9*(9). https://doi.org/10.3390/children9091407

Alfaraidy, H. A. (2020). Factors influencing Saudi parents' choice of International schools in Saudi Arabia. *Journal of Research in International Education*, *19*(3), 220–232. https://doi.org/10.1177/1475240920976259

Algouzi, S. (2014). *Discourse markers in Saudi English and British English: A comparative investigation of the use of English discourse markers.* Unpublished PhD Thesis, University of Salford. http://usir.salford.ac.uk/id/eprint/34008/

Algouzi, S. (2021). Functions of the discourse marker *so* in the LINDSEI-AR corpus. *Cogent Arts & Humanities*, *8*(1), 1872166. https://doi.org/10.1080/23311983.2021.1872166

Al-Harahsheh, A. M. & Kanakri, M. (2013). The Pragmatic functions and the translatability of "Ťayib" in Jordanian spoken Arabic. *US-China Foreign Language*, *11*(3). https://doi.org/10.17265/1539-8080/2013.03.004

Alhindawy, S. A. (2016). *A Longitudinal investigation of a sequential Arabic-English bilingual child's vocabulary development*. Unpublished MA thesis, Missouri State University. https://bearworks.missouristate.edu/theses/1887

Aljutaily, M. (2021). Prosodic correlates and pragmatic functions of the particle *tayeb* in spoken Saudi Arabic. *Journal of Language and Linguistic Studies*, *17*(4), 1932–1955. https://doi.org/Doi:10.52462/jlls.140

Allwood, J. (2003). Meaning potentials and context: Some consequences for the analysis of variation in meaning. In H. Cuyckens, R. Dirven, & J. R. Taylor (Eds), *Cognitive Approaches to Lexical Semantics* (pp. 29–66). Mouton De Gruyter. https://doi.org/10.1515/9783110219074.29

Al-Qahtani, Z., Wahed, A., & Al Zumor, A. (2016). Saudi parents' attitudes towards using English as a medium of instruction in private primary schools. *International Journal of Applied Linguistics and English Literature, 5 (1)*. 18–32. https://doi.org/10.7575/aiac.ijalel.v.5n.1p.18

Al-Yaari, S. A. S., Hammadi, F. S. A., Alyami, S., & Almaflehi, N. S. (2013). Using English discourse markers (EDMs) by Saudi EFL learners: A descriptive approach. *International Journal of English Language Education*, *1*(2), 1–26. https://doi.org/10.5296/ijele.v1i2.3058

Amador Moreno, C. P., McCarthy, M., & O'Keeffe, A. (2013). Can English provide a framework for Spanish response tokens? In J. Romero-Trillo (Ed), *The Yearbook of Pragmatics and Corpus Linguistics*, (pp. 175–201). Springer.

Ament, J. & Parés, J. B. (2018). The acquisition of discourse markers in the English-medium instruction context. In C. Pérez Vidal, S. López-Serrano, J. Ament, & D. J. Thomas-Wilhelm (Eds), *Learning context effects: Study abroad, formal instruction and international immersion classrooms* (pp. 43–74). Language Science Press.

Andersen. (2001). *Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents*. John Benjamins.

Andersen. (2015). Relevance. In K. Aijmer & C. Rühlemann (Eds), *Corpus pragmatics: A handbook* (pp. 169–194). Cambridge University Press.

Andersen, Brizuela, M., Dupuy, B., & Gonnerman, L. (1999). Cross-linguistic evidence for the early acquisition of discourse markers as register variables. *Discourse Markers in Language Acquisition, 31*(10). https://doi.org/10.1016/S0378-2166(98)00108-8

Andersen, G. (1998). *Are tag questions questions? Evidence from spoken data*. 19th ICAME Conference, Belfast.

Andersen, G. (2011). Corpus-based pragmatics I: qualitative studies. In W. Bublitz & N. R. Norrick (Eds), *Foundations of pragmatics*. Mouton de Gruyter.

Anthony, L. (2023). *AntConc (Version 4.2.4) [Computer Software]*. Waseda University. https://www.laurenceanthony.net/software

Arya, T. (2022). Exploring discourse marker use in Thai university students' conversations. *LEARN Journal: Language Education and Acquisition Research Network, 13*(1), 247–267. https://www.tci-thaijo.org/index.php/learn

Asik, A. & Cephe, P. T. (2013). Discourse markers and spoken English: Nonnative use in the Turkish EFL setting. *English Language Teaching*, *6*(12), 144–155. https://doi.org/10.5539/elt.v6n12p144

Aston, G. & Burnard, L. (1998). *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Baines, E. & Howe, C. (2010). Discourse topic management and discussion skills in middle childhood: The effects of age and task. *First Language*, *30*(3–4), 508–534. https://doi.org/10.1177/0142723710370538

Bakeer, S. (2023). *Discourse markers in doctoral supervision sessions: A multimodal perspective*. Routledge.

Baker, P. (2023). *Using corpora in discourse analysis* (Second edition). Bloomsbury.

Barbieri, M., Colavita, F., & Scheuer, N. (1989). Explanations: A pragmatic basis for early child competence. *IPrA Papers in Pragmatics.*, *3*(1), 129–154. https://doi.org/10.1075/iprapip.3.1.04bar

Bardovi-Harlig, K. & Bastos, M.-T. (2011). Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics*, *8*(3), 347-384. https://doi.org/10.1515/iprg.2011.017

Beach, W. A. (1993). Transitional regularities for 'casual' "Okay" usages. *Journal of Pragmatics*, *19*(4), 325–352. https://doi.org/10.1016/0378-2166(93)90092-4

Beach, W. A. (2020). Using prosodically marked "okays" to display epistemic stances and incongruous actions. *Journal of Pragmatics*, *169*, 151–164. https://doi.org/10.1016/j.pragma.2020.08.019

Beeching, K. (2015). Variability in native and non-native use of pragmatic markers: The example of *well* in role-play data. In K. Beeching & H. Woodfield (Eds), *Researching sociopragmatic variability: Perspectives from variational, interlanguage and contrastive pragmatics*. Palgrave Macmillan.

Beeching, K. (2016). *Pragmatic markers in British English: Meaning in social interaction*. Cambridge University Press.

Beeching, K., & Detges, U. (2014). Introduction. In K. Beeching & U. Detges (Eds.), *Discourse functions at the left and right periphery: Crosslinguistic investigations of language use and language change*. Brill.

Behrens, H. (Ed). (2008). *Corpora in language acquisition research: History, methods, perspectives*. John Benjamins.

Biber, D. (1993). Representativeness in corpus design, *Literary and Linguistic Computing*, *8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. John Benjamins.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (Eds). (1999). *Longman grammar of spoken and written English*. Pearson.

Bidgood, A., Pine, J. M., Rowland, C. F., & Ambridge, B. (2020). Syntactic representations are both abstract and semantically constrained: Evidence from children's and adults' comprehension and production/priming of the English passive. *Cognitive Science*, 44(9). https://doi.10.1111/cogs.12892

Birdsong, D. (2018). Plasticity, variability and age in L2 acquisition and bilingualism. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00081

Blakemore, D. (1987). *Semantic constraints on relevance*. Blackwell.

Blakemore, D. (1992). *Understanding utterances: An introduction to pragmatics*. Blackwell.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge University Press.

Blakemore, D. (2006). Discourse markers. In L. R. Horn (Ed), *the handbook of pragmatics* (pp. 221–240). Blackwell.

Blom, E., & Bosma, E. (2016). The sooner the better? An investigation into the role of age of onset and its relation with transfer and exposure in bilingual Frisian–Dutch children. *Journal of Child Language*, *43*(3), 581–607. https://doi.org/10.1017/S0305000915000574

Blom, E., Paradis, J. and Duncan, T.S. (2012), Effects of input properties, vocabulary size, and L1 on the development of third person singular –s in child L2 English. *Language learning*, 62: 965-994. https://doi.org/10.1111/j.1467-9922.2012.00715.x

Blum-Kulka, S., & Snow, C. E. (2004). Introduction: The potential of peer talk. *Discourse Studies*, *6*(3), 291-306. https://doi.org/10.1177/1461445604044290

Boersma, P., & Weenink, D. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Bolden, G. B. (2006). Little words that matter: Discourse markers "so" and "oh" and the doing of other-attentiveness in social interaction. *Journal of Communication*, *56*(4), 661–688. https://doi.org/10.1111/j.1460-2466.2006.00314.x

Bolden, G. B. (2008). "So what's up?": Using the discourse marker so to launch conversational business. *Research on Language and Social Interaction*, *41*(3), 302–337. https://doi.org/10.1080/08351810802237909

Bolden, G. B. (2009). Implementing incipient actions: The discourse marker 'so' in English conversation. *Pragmatic Markers*, *41*(5), 974–998. https://doi.org/10.1016/j.pragma.2008.10.004

Branigan, H. P., & McLean, J. F. (2016). What children learn from adults' utterances: An ephemeral lexical boost and persistent syntactic priming in adult-child dialogue. *Journal of Memory and Language*, 91, 141–157. https://doi.10.1016/j.jml.2016.02.002

Brezina, V. (2019). *Statistics in corpus linguistics*. Cambridge University Press.

Brezina, V. & McEnery, T. (2020). Introduction to corpus linguistics. In M. Paquot & N. Tracy-Ventura (Eds), *The Routledge handbook of L2 acquisition and corpora* (pp. 10–21). Routledge.

Brinton, L. J. (1990). The development of discourse markers. In J. Fisiak (Ed), *Historical linguistics and philology* (pp. 45–72). Mouton de Gruyter.

Brinton, L. J. (1996). *Pragmatic markers in English: Grammaticalisation and discourse functions*. Mouton de Gruyter.

Brinton, L. J. (2006). Pathways in the development of pragmatic markers in English. In A. van Kemenade & B. Los (Eds), *The handbook of the history of English* (pp. 307–334). Blackwell.

Brinton, L. J. (2007). The Development of *I mean*: Implications for the study of historical pragmatics. In S. M. Fitzmaurice & I. Taavitsainen (Eds), *Methods in Historical Pragmatics* (pp. 37–80). Mouton de Gruyter.

Brinton, L. J. (2017). *The evolution of pragmatic markers in English: Pathways of change*. Cambridge University Press.

Brinton, L. J. (2023). *Pragmatics in the history of English*. Cambridge University Press.

Bryant, J. B. (2009). Pragmatic development. In E. L. Bavin (Ed), *The Cambridge handbook of child language*. Cambridge University Press.

Bu, J. (2013). A study of the acquisition of discourse markers by Chinese learners of English. *International Journal of English Studies*, *13*(1), 29–50. https://doi.org/10.6018/ijes/2013/1/152681

Buchstaller, I. (2001). An Alternative View of 'like': Its grammaticalisation in conversational American English and beyond. *Edinburgh Working Papers in Applied Linguistics*, *11*, 21–41.

Buchstaller, I. (Ed). (2014). *Quotatives: New trends and sociolinguistic implications*. Wiley-Blackwell.

Buysse, L. (2007). Discourse marker *so* in the English of Flemish university students. *BELL : Belgian Journal of English Language and Literatures*, *5*, 79–95. http://hdl.handle.net/1854/LU-2973995

Buysse, L. (2009). 'So' as a marker of elaboration in native and non-native speech. In S. Slembrouck, M. Taverniers, & M. Van Herreweghe (Eds), *From will to well: Studies in linguistics offered to Anne-Marie Simon-Vandenbergen* (pp. 79–91). Academia Press.

Buysse, L. (2010). Discourse markers in the English of Flemish university students. In Witczak-Plisiecka, I. (Ed), *Pragmatic perspectives on language and linguistics: Speech Actions in Theory and Applied Studies* (pp. 461–484). Cambridge Scholars Publishing.

Buysse, L. (2012). *So* as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics, 44*(13), 1764–1782. https://doi.org/10.1016/j.pragma.2012.08.012

Buysse, L. (2014). "*So* what's a year in a lifetime *so*." Non-prefatory use of *so* in native and learner English. *Text & Talk-An Interdisciplinary Journal of Language Discourse Communication*, *34*(1), 23–47. https://doi.org/10.1515/text-2013-0036

Buysse, L. (2015). 'Well it's not very ideal …' The pragmatic marker *well* in learner English. *Intercultural Pragmatics*, *12*(1), 59–89. https://doi.org/doi:10.1515/ip-2015-0003

Buysse, L., & Blanchard, M. (2022). L1 and non-L1 perceptions of discourse markers in English. *Pragmatics & Cognition*, *29*(2), 222–245. https://doi.org/10.1075/pc.21023.bla

Bybee, J. L. (2006). *Frequency of use and the organization of language*. Oxford University Press.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press https://doi.org/10.1017/CBO9780511750526

Caldwell-Harris, C. L., & MacWhinney, B. (2023). Age effects in L2 acquisition: Expanding the emergentist account. *Brain and Language*, *241*(2), 105–269. https://doi.org/10.1016/j.bandl.2023.105269

Caliendo, G., & Magliacane, A. (2020). Sociopragmatic competence and L2 acquisition: Learners of English in a study abroad context. In G. Caliendo, R. Janssens, S. Slembrouck, & P. V. Avermaet (Eds), *Bridging the Gap between Language Policies and Language Practices* (pp. 167–190). Mouton De Gruyter. https://doi.org/doi:10.1515/9781501503207-008

Carlson, L. (1984). *'Well' in dialogue games: A discourse analysis of the interjection well in idealized conversation*. John Benjamins.

Carter, R. & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide; spoken and written English grammar and usage*. Cambridge University Press.

Castro, C. (2009). The use and functions of discourse markers in EFL classroom interaction. *Profile Issues in teachers' professional development*, *11*, 57–77.

Cekaite, A. (2007). A Child's development of interactional competence in a Swedish L2 classroom. *The Modern Language Journal*, *91*(1), 45–62. https://doi.org/10.1111/j.1540-4781.2007.00509.x

Cekaite, A. & Aronsson, K. (2014). Language play, peer group improvisations, and L2 learning. In A. Cekaite, S. Blum-Kulka, V. Grøver, & E. Teubal (Eds), *Children's peer talk: Learning from each other* (pp. 194–213). Cambridge University Press.

Cekaite, A., Blum-Kulka, S., Grøver, V., & Teubal, E. (2014). Children's peer talk & learning: uniting discursive, social, and cultural facets of peer interactions: editors' introduction. In A. Cekaite, S. Blum-Kulka, V. Grøver, & E. Teubal (Eds), *Children's peer talk: Learning from each other* (pp. 3–19). Cambridge University Press.

Celle, A., & Huart, R. (2007). Connectives as discourse landmarks: Introduction. In A. Celle & R. Huart (Eds), *Connectives as discourse landmarks*. John Benjamins.

Chee, M., Jones, F., Morford, J & Shin, N. (2023). Usage-based approaches to child language development. In M. Díaz-Campos & S. Balasch (Eds.), *The handbook of usage-based linguistics* (pp. 379-392). John Wiley & Sons, Inc.

Chen, J. (2019). What are discourse markers? *In Proceedings of the 2nd Symposium on Health and Education 2019 (SOHE 2019)*, (pp1–9). https://doi.org/10.2991/sohe-19.2019.1

Chomsky, N. (1965). Aspects of the theory of syntax. M.I.T. Press.

Colletta, J.-M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., & Levine, S. (2015). Effects of age and language on co-speech gesture production: An investigation of French, American, and Italian children's narratives. *Journal of Child Language*, *42*(1), 122–145. https://doi.org/10.1017/S0305000913000585

Collier, V. P. (1987). *The effect of age on acquisition of a L2 for school*. National Clearinghouse for Bilingual Education.

Collins, L. (2019). *Corpus linguistics for online communication: A guide for research*. Routledge.

Condon, S. L. (1986). The discourse functions of *OK*. *Semiotica*, *60*(2), 73–102. https://doi.org/doi:10.1515/semi.1986.60.1-2.73

Cook-Gumperz, J. & Kyratzis, A. (2015). Child discourse. In D. Tannen, H. Hamilton, & D. Schiffrin (Eds), *The Handbook of Discourse Analysis* (pp. 681–704). https://doi.org/10.1002/9781118584194.ch32

Corrigan, K. P. (2015). "I always think of people here, you know, saying 'like' after every sentence". In C. P. A. Moreno, K. McCafferty, & E. Vaughan (Eds), *Pragmatic markers in Irish English* (pp. 37–64). John Benjamins.

Crible, L. (2017). Discourse markers and (dis)fluency in English and French: Variation and combination in the DisFrEn corpus. *International Journal of Corpus Linguistics*, *22*(2), 242–269. https://doi.org/10.1075/ijcl.22.2.04cri

Crible, L. (2018). *Discourse markers and (dis)fluency: Forms and functions across languages and registers*. John Benjamins.

Crible, L. & Blackwell, S. E. (2020). Introduction: Discourse-pragmatic markers in speech and sign. *Journal of Pragmatics*, *156*, 24–27. https://doi.org/10.1016/j.pragma.2019.09.002

Crible, L. & Degand, L. (2019). Domains and functions: A two-dimensional account of discourse markers. *Discours*, *24*. https://doi.org/10.4000/discours.9997

Crible, L. & Pascual, E. (2020). Combinations of discourse markers with repairs and repetitions in English, French and Spanish. *Journal of Pragmatics*, *156*, 54–67. https://doi.org/10.1016/j.pragma.2019.05.002

Crible, L. & Zufferey, S. (2015). Using a unified taxonomy to annotate discourse markers in speech and writing. *In Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*, (pp14–22). https://aclanthology.org/W15-0202

Culpeper, J., Mackey, A., & Taguchi, N. (2018). *Second language pragmatics: From theory to research*. Routledge.

Dailey-O'Cain, J. (2000). The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*. *Journal of Sociolinguistics, 4 (1)*, 60–80. https://doi.org/10.1111/1467-9481.00103

D'Arcy, A. (2006). Lexical replacement and the like(s). *American Speech*, *81*, 339–357. https://doi.org/10.1215/00031283-2006-024

D'Arcy, A. (2007). *Like* and language ideology: disentangling fact from fiction. *American Speech*, *82*(4). https://doi.org/10.1215/00031283-2007-025

D'Arcy, A. (2017). *Discourse-pragmatic variation in context: Eight hundred years of* like. John Benjamins.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14* (2), 159–190. John Benjamins. https://doi.org/10.1075/ijcl.14.2.02dav

Davies, M. (2015). Corpora: An introduction. In D. Biber & R. Reppen (Eds), *The Cambridge handbook of English corpus linguistics*. Cambridge University Press.

Davies, M. (2019). Corpus-based studies of lexical and semantic variation: The importance of both corpus size and corpus design. In C. Suhr, T. Nevalainen, & I. Taavitsainen (Eds), *From data to evidence in English language research* (pp. 66–87). Brill.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *BELL : Belgian Journal of English Language and Literatures*, 225–246.

De Houwer, A. (2021). *Bilingual development in childhood*. Cambridge University Press.

Degand, L. & Simon, A. C. (2009). On identifying basic discourse units in speech: Theoretical and empirical issues. *Discours*, (4). https://doi.org/10.4000/discours.5852

Demuth, K. (2008). Exploiting corpora for language acquisition research. In H. Behrens (Ed), *Corpora in language acquisition research: History, methods, perspectives*. John Benjamins.

Dinkar, T., Colombo, P., Labeau, M., & Clavel, C. (2020). The importance of fillers for text representations of speech transcripts. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 7985–7993). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.641

Diskin, C. (2017). The use of the discourse-pragmatic marker '*like*' by native and non-native speakers of English in Ireland. *Journal of Pragmatics*, *120*, 144–157. https://doi.org/10.1016/j.pragma.2017.08.004

Diskin-Holdaway, C. (2021). *You know* and *like* among migrants in Ireland and Australia. *World Englishes*, *40*(4), 562–577. https://doi.org/10.1111/weng.12541

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in L2 acquisition*. Lawrence Erlbaum.

Dörnyei, Z. & Skehan, P. (2003). Individual differences in L2 learning. In C. J. Doghty & M. H. Long (Eds), *the Handbook of Second Language Acquisition* (pp. 589–630). Blackwell Publishing. https://doi.org/10.1002/9780470756492.ch18

Drummond, K. & Hopper, R. (1993a). Acknowledgment tokens in series. *Communication Reports*, *6*(1), 47–53. https://doi.org/10.1080/08934219309367561

Drummond, K. & Hopper, R. (1993b). Back channels revisited: acknowledgment tokens and speakership incipiency. *Research on Language and Social Interaction*, *26*, 157–177. https://doi.org/10.1207/s15327973rlsi2602_3

Drummond, K. & Hopper, R. (1993c). Some uses of *yeah*. *Research on language and social interaction*, *26*(2), 203–212. https://doi.org/10.1207/s15327973rlsi2602_6

Duncan, S. & Donald, F. (1977). *Face-to-face interaction research, methods, and theory*. Lawrence Erlbaum.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61–74.

Ehrman, M. E., Leaver, B. L., & Oxford, R. L. (2003). A brief overview of individual differences in second language learning. *System*, *31*(3), 313–330. https://doi.org/10.1016/S0346-251X(03)00045-9

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, *24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, R. (1994). *The study of L2 acquisition*. Oxford University Press.

Ellis, R. (2015). *Understanding L2 acquisition*. Routledge, Taylor & Francis Group.

Erman, B. (1987). *Pragmatic expressions in English: A study of 'You know, you see' and 'I mean' in face-to-face conversation*. Almqvist & Wiksell.

Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics*, *33*, 1337–1359. https://doi.org/10.1016/S0378-2166(00)00066-7

Escalera, E. A. (2002). *The role of context in children's use of discourse markers: Three studies of peer conversation*. Unpublished PhD Thesis, University of California, Berkeley.

Escalera, E. A. (2009). Gender differences in children's use of discourse markers: Separate worlds or different contexts? *Journal of pragmatics*, *41*(12). 2479–2495 https://doi.org/10.1016/j.pragma.2006.08.013

Fernández, J. & Staples, S. (2020). Pragmatic approaches. In M. Paquot & N. Tracy-Ventura (Eds), *The Routledge handbook of L2 acquisition and corpora* (pp. 239–251). Routledge.

Fischer, K. (2000). Discourse particles, turn-taking, and the semantics-pragmatics interface. *Revue de Sémantique et Pragmatique*, *8*, 111–137.

Fischer, K. (2006). Towards an understanding of the spectrum of approaches to discourse particles: Introduction to the volume. In K. Fischer (Eds), *Approaches to discourse particles* (pp. 1–20). Elsevier.

Fischer, K. (2014). Discourse markers. In K. P. Schneider & A. Barron (Eds), *Pragmatics of discourse* (pp. 271–294). Mouton De Gruyter.

Flores, C., Santos, A. L., & Marques, R. (2017). Age and input effects in the acquisition of mood in heritage Portuguese. *Journal of Child Language*, *44*(4), 795–828. https://doi.org/doi:10.1017/S0305000916000222

Forchini, P. (2010). "*I mean*, what was that about?"*. Spontaneous and movie conversation compared. *L'analisi Linguistica e Letteraria*, *XVIII*(2), 323–334.

Fox Tree, J. E. (2006). Placing *like* in telling stories. *Discourse Studies*, *8*(6), 723–743. https://doi.org/10.1177/1461445606069287

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an *oh* makes. *Journal of Memory and Language*, *40*(2), 280–295. https://doi.org/10.1006/jmla.1998.2613

Fox Tree, J. E., & Schrock, J. C. (2002). Basic meanings of *you know* and *I mean*. *Journal of Pragmatics*, *34*(6), 727–747. https://doi.org/10.1016/S0378-2166(02)00027-9

Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, *38*, (1/4), 19–33. http://www.jstor.org/stable/44362602.

Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics*, *14*(3), 383-398. https://doi.org/10.1016/0378-2166(90)90096-V

Fraser, B. (1996). Pragmatic markers. *Pragmatics*, *6*(2), 167–190. https://doi.org/10.1075/prag.6.2.03fra

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*, 931–952. https://doi.org/10.1016/S0378-2166(98)00101-5

Fraser, B. (2006). Towards a theory of discourse markers. In K. Fischer (Ed), *Approaches to discourse particles* (pp. 189–204). Elsevier.

Fraser, B. (2009). An account of discourse markers. *International Review of Pragmatics*, *1*(2), 293–320. https://doi.org/10.1163/187730909X12538045489818

Fraser, B. (2021). An introduction to discourse markers. In J. C. Félix-Brasdefer & R. Shively (Eds), *New directions in L2 pragmatics* (pp. 314–335). Mouton De Gruyter.

Fuller, J. (2003a). Discourse marker use across speech contexts: A comparison of native and non-native speaker performance. *Multilingua: Journal of Cross-Cultural and Interlanguage Communication*, *22*, 185–208. https://doi.org/10.1515/mult.2003.010

Fuller, J. (2003b). The influence of speaker roles on discourse marker use. *Journal of Pragmatics*, *35*(1), 23–45. https://doi.org/10.1016/S0378-2166(02)00065-6

Fuller, J. (2003c). Use of the discourse marker *like* in interviews. *Journal of Sociolinguistics*, *7*(3), 365–377. https://doi.org/10.1111/1467-9481.00229

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, *28*(3), 410–439. https://doi.org/10.1093/applin/amm030

Furkó, P. B. (2009). *The Pragmatic Marker-discourse Marker Dichotomy Reconsidered: The Case of* Well *and* Of Course. Unpublished PhD Thesis, University of Debrecen.

Furkó, P. B. (2020). *Discourse markers and beyond: Descriptive and critical perspectives on discourse-pragmatic devices across genres and languages*. Palgrave Macmillan.

Gablasova, D., Brezina, V., Mcenery, T., & Boyd, E. (2017). Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics*, *38*(5), 613–637. https://doi.org/10.1093/applin/amv055

Gaines, P. (2011). The multifunctionality of discourse operator *okay*: evidence from a police interview. *Journal of Pragmatics*, *43*(14), 3291–3315. https://doi.org/10.1016/j.pragma.2011.06.005

Gardner, R. (1998). Between speaking and listening: the vocalisation of understandings. *Applied Linguistics*, *19*, 204–224. https://doi.org/10.1093/applin/19.2.204

Gerard, J., Keller, F., & Palpanas, T. (2010). Corpus Evidence for age effects on priming in child language. In Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 218-223) https://mindmodeling.org/cogsci2010/papers/0036/index.html

Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference? In J. Romero-Trillo (Ed), *A mutualistic entente* (pp. 119–150). Mouton De Gruyter. https://doi.org/doi:10.1515/9783110199024.119

Gilquin, G. (2016). Discourse markers in L2 English: From classroom to naturalistic input. In O. Timofeeva, A.-C. Gardner, A. Honkapohja, & S. Chevalier (Eds), *New approaches to English linguistics: Building bridges* (pp. 213–249). John Benjamins.

Gilquin, G., De Cock, S., & Granger, S. (2010). *The Louvain International Database of Spoken English Interlanguage.* Handbook and CD-ROM. Louvain-la-Neuve: Presses universitaires de Louvain.

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Goodwin, M. H. & Kyratzis, A. (2007). Children socializing children: practices for negotiating the social order among peers. *Research on Language and Social Interaction*, *40*(4), 279–289. https://doi.org/10.1080/08351810701471260

Granger, S. (2002). A Bird's-eye view of learner corpus research. In S. Granger (Ed), *Computer learner corpora, L2 acquisition and foreign language teaching* (pp. 3-33). John Benjamins.

Granger, S. (2009). The contribution of learner corpora to L2 acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed), *Studies in Corpus Linguistics* (pp. 13–332). John Benjamins.

Granger, S. (2021). Have learner corpus research and L2 acquistion finally met? In B. L. Bruyn & M. Paquot (Eds), *Learner corpus research meets L2 acquisition* (pp. 243–257). Cambridge University Press.

Granger, S., Gilquin, G., & Meunier, F. (Eds). (2015). *The Cambridge handbook of learner corpus research*. Cambridge University Press.

Greenbaum, S. & Nelson, G. (1996). The International Corpus of English (ICE) Project. *World Englishes*, *15*(1), 3–15. https://doi.org/10.1111/j.1467-971X.1996.tb00088.x

Gries, S. (2015). Quantitative designs and statistical techniques. In D. Biber & R. Reppen (Eds), *The Cambridge handbook of English corpus linguistics* (pp. 50-72). Cambridge University Press.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15. https://doi.org/10.1016/0093-934x(89)90048-5

Gui, S. C. & Yang, H. Z. (2001). *Chinese learner English corpus.* Shanghai Foreign Language Education Press.

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. Routledge, Taylor & Francis Group.

Hardie, A. (2012). CQPweb—Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, *17*(3), 380–409. https://doi.org/10.1075/ijcl.17.3.04har

Hardie, A. (2016). Corpus linguistics. In K. Allan (Ed), *The Routledge handbook of linguistics* (pp. 502–515). Routledge.

Haselow, A. (2017). *Spontaneous spoken English: An integrated approach to the emergent grammar of speech*. Cambridge University Press.

Haselow, A. (2021). The acquisition of pragmatic markers in the foreign language classroom: An experimental study on the effects of implicit and explicit learning. *Journal of Pragmatics*, *186*, 73–86. https://doi.org/10.1016/j.pragma.2021.09.017

Haznedar, B. & Gavruseva, E. (Eds). (2008). *Current trends in child L2 acquisition: A generative perspective*. John Benjamins.

Haznedar, B., & Gavruseva, E. (2008). Recent perspectives in child second language acquisition. In B. Haznedar & E. Gavruseva (Eds.), *Current trends in child second language acquisition: A generative perspective*. John Benjamins.

Hellermann, J. & Vergun, A. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Focus-on Issue: Topics in Applied Pragmatics*, *39*(1), 157–179. https://doi.org/10.1016/j.pragma.2006.04.008

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds), *Structures of Social Action: Studies in Conversation Analysis* (pp. 299–345). Cambridge University Press.

Heritage, J. (1998). *Oh*-prefaced responses to inquiry. *Language in Society*, *27*(3), 291–334. Cambridge Core. https://doi.org/10.1017/S0047404500019990

Heritage, J. (2015). *Well*-prefaced turns in English conversation: A conversation analytic perspective. *Journal of Pragmatics*, *88*, 88–104. https://doi.org/10.1016/j.pragma.2015.08.008

Heritage, J. (2016). On the diversity of changes of state and their indices. *Journal of Pragmatics*, *104*, 207–210. https://doi.org/10.1016/j.pragma.2016.09.007

Heritage, J. (2018). Turn-initial particles in English: The cases of *oh* and *well*. In J. Heritage & M.-L. Sorjonen (Eds), *Between turn and sequence: Turn-initial particles across languages* (pp. 155–190). John Benjamins.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1). https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E. (2014). *Language development* (Fifth edition). Wadsworth Cengage Learning.

Holmes, J. (1986). Functions of *you know* in women's and men's speech. *Language in Society*, *15*(1), 1–21. https://doi.org/10.1017/S0047404500011623

House, J. (1996). Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition*, *18*(2), 225–252. https://doi.org/10.1017/S0272263100014893

House, J. (2013). Developing pragmatic competence in English as a lingua franca: Using discourse markers to express (inter)-subjectivity and connectivity. *Journal of Pragmatics*, *59*(A), 57-67. https://doi.org/10.1016/j.pragma.2013.03.001

Huang, L. (2011). *Discourse markers in spoken English: A corpus study of native speakers and Chinese non-native speakers*, Unpublished PhD Thesis, University of Birmingham.

Huang, L. (2019). A corpus-based exploration of the discourse marker *well* in spoken interlanguage. *Language and Speech*, *62*(3), 570–593. https://doi.org/10.1177/0023830918798863

Huang, L., Lin, Y., & Gráf, T. (2023). Development of the use of discourse markers across different fluency levels of CEFR. *Pragmatics*, *33*(1), 49–77. https://doi.org/10.1075/prag.21016.hua

Huddlestone, K. & Fairhurst, M. (2013). The pragmatic markers *anyway*, *okay*, and *shame*: A South African English corpus study. *Stellenbosch Papers in Linguistics Plus*, *24*(1), 93–110. https://doi.org/10.5842/42-0-170

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Huo, H. (2023). A corpus-based study on discourse markers in native and non-native spoken English. In L. F. Ying, Ts. Dr. S. H. Halili, & D. Mishra (Eds), *Proceedings of the 2nd International Conference on Education, Language and Art (ICELA 2022)* (pp. 167–176). Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-004-6_23

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language*, 50(2), 182–195. https://doi.10.1016/j.jml.2003.09.003

Ingram, D. (1998). Assessment of Language Proficiency. *PASAA*, *28*(1), 1–25. https://doi.org/10.58837/CHULA.PASAA.28.1.1

Ismail, S. & Shaban, M. (2016). Prospective teachers' views toward the use of English as the medium of instruction in early childhood classes in the United Arab Emirates. *International Journal of Research Studies in Language Learning*, *5*, 117-128. https://doi.org/10.5861/ijrsll.2016.1360

Jakupčević, E. (2019). Young language learners' use of discourse markers in L2 narratives. *English Teaching & Learning*, *43*(4), 411-428. https://doi.org/10.1007/s42321-019-00039-w

James, D. (1978). The use of *oh*, *ah*, *say*, and *well* in relation to a number of grammatical phenomena. *Paper in Linguistics*, *11*(3-4), 517–535. https://doi.org/10.1080/08351817809389186

Jarrah, A. (2020). Elementary schools mathematics and science teachers' perspectives on using English as a medium of instruction. *Humanities & Social Sciences Reviews*, *8*(4), 473–482. https://doi.org/10.18510/hssr.2020.8446

Jefferson, G. (1983). Notes on a systematic deployment of the acknowledgement tokens "*yeah*"; and "*mm hm*"; *Paper in Linguistics*, *17*(2), 197-216. https://doi.org/10.1080/08351818409389201

Jucker, A. H. (1993). The discourse marker *well*: A relevance-theoretical account. *Journal of Pragmatics*, *19*(5), 435–452. https://doi.org/10.1016/0378-2166(93)90004-9

Jucker, A. H. (1997). The discourse marker *well* in the history of English. *English Language and Linguistics*, *1*(1), 91-110. https://doi.org/10.1017/S136067430000037X

Jucker, A. H., & Smith, S. W. (1998). And people just *you know* like 'wow': Discourse markers as negotiating strategies. In A. H. Jucker & Y. Ziv (Eds), *Discourse markers: Descriptions and theory* (pp. 171-202). John Benjamins.

Jucker, A. H., & Ziv, Y. (Eds). (1998). *Discourse markers: Descriptions and theory*. John Benjamins.

Jun, SA., & Fougeron, C. (2001). A Phonological Model of French Intonation. In: Botinis, A. (Eds.) *intonation. text, speech and language technology*. Springer.

Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.

Kerswill, P. (1996). Children, adolescents, and language change. *Language Variation and Change*, *8*(2), 177–202. https://doi.org/10.1017/S0954394500001137

Kerswill, P., Cheshire, J., Fox, S., & Torgersen, E. (2013). English as a contact language: The role of children and adolescents. In D. Schreier & M. Hundt (Eds), *English as a Contact Language* (pp. 258–282). Cambridge University Press. https://doi.org/10.1017/CBO9780511740060.015

Kidd, E. (2012). Individual differences in syntactic priming in language acquisition. *Applied Psycholinguistics*, 33(2), 393–418. https://doi.10.1017/S0142716411000415

Kirkpatrick, A. (2010). Researching English as a lingua franca in Asia: the Asian Corpus of English (ACE) project. *Asian Englishes*, 13(1), 4–18. https://doi.org/10.1080/13488678.2010.10801269

Kim, I. L. (2011). *Development of discourse markers in Korean and English monolingual children and in Korean-English bilingual children*. Unpublished PhD Thesis, Temple University.

Klein, C. (2014, March 21). *The birth of* ok. History.com. Retrieved November 16, 2024, from https://www.history.com/news/the-birth-of-ok-175-years-ago

Koops, C., & Lohmann, A. (2022). Explaining reversible discourse marker sequences: A case study of *and* and *so*. *Journal of Pragmatics*, *191*, 156–171. https://doi.org/10.1016/j.pragma.2022.01.014

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Pergamon.

Kumarage, S., Donnelly, S., & Kidd, E. (2022). Implicit learning of structure across time: A longitudinal investigation of syntactic priming in young English-acquiring children. *Journal of Memory and Language*, 127, 1–26. https://doi.10.1016/j.jml.2022.104374

Kurdi, H. (2008). *The use of discourse markers by Syrian Arabic learners of English*. Unpublished PhD Thesis, University of Manchester.

Kwon, Y. (2022). Pragmatic discourse markers *you know* and *sort of* in Korean EFL teacher corpus. *Studies in Foreign Language Education*, *34*, 291–322. https://doi.org/10.16933/sfle.2020.34.1.291

Kyratzis, A. (2004). Talk and Interaction among children and the co-construction of peer groups and peer culture. *Annual Review of Anthropology*, *33*, 625–649. https://doi.org/10.1146/annurev.anthro.33.070203.144008

Kyratzis, A., & Ervin-Tripp, S. (1999). The development of discourse markers in peer interactions. *Journal of Pragmatics*, *31*, 1321–1338. https://doi.org/10.1016/S0378-2166(98)00107-6

Labov, W. (2001). *Principles of linguistic change. 2: Social factors*. Wiley-Blackwell.

Lakoff, R. (1973). Questionable answers and answerable questions. In B. B. Kachru, R. B. Lees, Y. Malkiel, A. Pietrangeli, & S. Saporta (Eds), *Issues in linguistics: Papers in honor of Henry and Renée Khane* (pp. 453–467). University of Illinois Press.

Lam, P. W. (2007). *Discourse particles in an intercultural corpus of spoken English*. Unpublished PhD Thesis. Hong Kong Polytechnic University.

Landis, J. R, & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 8-29). Longman.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991* (pp. 105-126). Mouton de Gruyter.

Leech, G. (1997). Introducing corpus annotation. In R. Garside, G. Leech, & T. McEnery (Eds), *Corpus annotation: Linguistic information from computer text corpora* (pp. 1-18). Longman.

Leech, G. (2005). Adding linguistic annotation. In M. Wynne (Ed), *Developing linguistic corpora: A guide to good practice* (pp. 17-29). Oxbow Books.

Leech, G. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In Hundt, M., Nesselhauf, N., & Biewer, C. (Eds), *Corpus linguistics and the web*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789401203791_009

Leech, G., Myers, G., & Thomas, J. (1995). Introduction. In G. Leech, G. Myers, & J. Thomas (Eds), *Spoken English on computer: Transcription, mark-up and application* (pp. 15–18). Longman.

Lenk, U. (1998). *Marking discourse coherence: Functions of discourse markers in spoken English*. G. Narr.

Lenneberg, E. H. (1967). *Biological foundations of language*. Wiley.

Lennon, P. (2009). Contrastive analysis, error analysis, interlanguage. In S. Gramley & V. Gramley (Eds.), *Bielefeld introduction to applied linguistics: A course book* (pp. 51–60). Aisthesis-Verl.

Levelt, W. The ability to speak: from intentions to spoken words. *European Review*, *3*(1), 13-23. https://doi:10.1017/S1062798700001290

Levey, S. (2006). The sociolinguistic distribution of discourse marker *like* in preadolescent speech. *Multilingua*, *25*(4), 413–441. https://doi.org/doi:10.1515/MULTI.2006.022

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Lewis, D. M. (2006). Discourse markers in English: A discourse-pragmatic view. In K. Fischer (Ed), *Approaches to discourse particles* (pp. 43-59). Elsevier.

Li, M, & Xiao, Y. (2012). A comparative study on the use of the discourse marker "*well*" by Chinese learners of English and native English speakers. *International Journal of English Linguistics*, *2*(5), 65–71. https://doi.org/10.5539/ijel.v2n5p65

Liao, S. (2009). Variation in the use of discourse markers by Chinese teaching assistants in the US. *Journal of Pragmatics*, *41*(7), 1313–1328. https://doi.org/10.1016/j.pragma.2008.09.026

Lightbown, P., & Spada, N. (2006). *How languages are learned*. Oxford University.

Liu, B. (2013). Effect of first language on the use of English discourse markers by L1 Chinese speakers of English. *Journal of Pragmatics*, *45*, 149–172. https://doi.org/10.1016/j.pragma.2012.11.002

Liu, B. (2016). Effect of L2 exposure: From a perspective of discourse markers. *Applied Linguistics Review*, *7*(1), 73-98. https://doi.org/10.1515/applirev-2016-0004

Llanes, À. & Muñoz, C. (2012). Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning*, *64*(1), 63-90. https://doi.org/10.1111/j.1467-9922.2012.00731.x

Love, R., Dembry, C., Hardie, A., Brezina, V. & McEnery, T. (2017). The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations, *International Journal of Corpus Linguistics*, *22*(3), 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

Macaulay, R. (2002). *You know*, it depends. *Journal of Pragmatics*, *34*(6), 749–767. https://doi.org/10.1016/S0378-2166(01)00005-4

Magliacane, A., & Howard, M. (2019). The role of learner status in the acquisition of pragmatic markers during study abroad: The use of '*like*' in L2 English. *Journal of Pragmatics*, *146*, 72–86. https://doi.org/10.1016/j.pragma.2019.01.026

Maschler, Y., & Schiffrin, D. (2015). Discourse markers language, meaning, and context. In Y. Maschler & D. Schiffrin (Eds.), the handbook of discourse analysis (pp. 189–221). John Wiley & Sons.

McCarthy, M. & O'Keeffe, A. (2010). Historical perspective: What are corpora and how have they evolved? In A. O'Keeffe & M. McCarthy (Eds), *The Routledge handbook of corpus linguistics* (pp. 3–13). Routledge.

McDonough, K., & Trofimovich, P. (2015). Structural priming and the acquistion of novel form-meaning mappings. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 105-124). De Gruyter Mouton.

McEnery, T. & Gabrielatos, C. (2006). English corpus linguistics. In B. Aarts & A. McMahon (Eds), *The Handbook of English Linguistics* (pp. 33–71). Blackwell.

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

McEnery, T. & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh University Press.

McEnery, T. Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

McEnery, T. Brezina, V., Gablasova, D., & Jayanti, B. (2019). Corpus linguistics, learner corpora, and SLA: employing technology to analyze language use. *Annual Review of Applied Linguistics*, *39*, 74–92. https://doi.org/doi:10.1017/S0267190519000096

Meehan, T. (1991). It's Like, 'What's Happening in the Evolution of Like?': A theory of grammaticalisation. *Kansas Working Papers in Linguistics*, *16*, 37-51. https://doi.org/10.17161/KWPL.1808.423

Mei, W. (2012). *A corpus-based comparative study of pragmatic markers:* I mean *and* you know *in native and non-native conversation*, Unpublished PhD Thesis, University of Liverpool.

Merritt, M. (1978). On the use of '*OK*' in service encounters. In R. Bauman & J. Sherzer (Eds), *Working papers in Sociolinguistics* (pp. 1–17). University of Texas.

Miller, J. (2009). *Like* and other discourse markers. In P. Peters, P. Collins, & A. Smith (Eds), *Varieties of English around the world* (pp. 315–336). John Benjamins.

Miller, J., & Weinert, R. (1995). The function of LIKE in dialogue. *Journal of Pragmatics*, *23*, 365–393. https://doi.org/10.1016/0378-2166(94)00044-F

Ministry of Education, KSA. (2021). General statistics. *Education Database*. Retrieved June 21, 2022, from https://data.gov.sa/Data/ar/organization/ministry_of_education

Montero, L., Serrano, R., & Llanes, A. (2017). The influence of learning context and age on the use of L2 communication strategies. *Language Learning Journal*, *45*(1), 1–16. https://doi.org/10.1080/09571736.2013.853824

Mosegaard Hansen, M.-B. (1998). *The function of discourse particles: A study with special reference to spoken standard French*. John Benjamins.

Mosegaard Hansen, M.-B. (2006). A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French *toujours*). In K. Fischer (Ed), *Approaches to discourse particles* (pp. 21-41). Elsevier.

Müller, S. (2004). 'Well you know that type of person': Functions of *well* in the speech of American and German students. *Journal of Pragmatics*, *36*(6), 1157–1182. https://doi.org/10.1016/j.pragma.2004.01.008

Müller, S. (2005). *Discourse markers in native and non-native English discourse*. John Benjamins.

Nicholls, D. (2003). The Cambridge Learner Corpus—Error coding and analysis for lexicography and ELT. In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds), *Proceedings of the 2003 Corpus Linguistics Conference* (pp. 572–581). Lancaster University.

Nippold, M. (2004). Research on later language development: International perspective. In R. A. Berman (Ed), *Language development across childhood and adolescence*, (pp1–8). John Benjamins.

Norén, K. & Linell, P. (2007). Meaning potentials and the interaction between lexis and contexts: An empirical substantiation. In *Pragmatics 17* (3), 387–416. John Benjamins.

Norrick, N. R. (2001). Discourse markers in oral narrative. *Journal of Pragmatics*, *33*(6), 849–878. https://doi.org/10.1016/S0378-2166(01)80032-1

Norrick, N. R. (2009). Interjections as pragmatic markers. *Journal of Pragmatics*, *41*(5), 866–891. https://doi.org/10.1016/j.pragma.2008.08.005

Odato, C. V. (2013). The development of children's use of discourse *like* in peer interaction. *American Speech*, *88*(2) 117–143. https://doi.org/10.1215/00031283-2346825

O'Keeffe, A. (2018). Corpus-based function-to-form approaches. In A. H. Jucker, K. P. Schneider, & W. Bublitz (Eds), *Methods in pragmatics* (pp. 587–618). Mouton De Gruyter.

O'Keeffe, A. & Adolphs, S. (2008). Response tokens in British and Irish discourse: Corpus, context and variational pragmatics. In K. P. Schneider & A. Barron (Eds), *Pragmatics & Beyond New Series*, (pp. 69–98). John Benjamins. https://doi.org/10.1075/pbns.178.05ok

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language use and language teaching.* Cambridge University Press.

Östman, J.-O. (1981). You know*: A discourse functional approach*. John Benjamins.

Östman, J.-O. (1982). *The Symbiotic Relationship between Pragmatic Particles and Impromptu Speech.* ERIC Clearinghouse

Othman, Z. (2007). *An exploratory study of discourse markers: Their distribution, meanings and uses in a corpus of academic lectures*. Unpublished PhD Thesis. Lancaster University.

Othman, Z. (2010). The use of *okay, right* and *yeah* in academic lectures by native speaker lecturers: Their 'anticipated' and 'real' meanings. *Discourse Studies*, *12*(5), 665–681. https://doi.org/10.1177/1461445610376365

Owen, M. (1981). Conversational Units and the Use of *Well...*'. In P. Werth (Ed), *Conversation and discourse* (pp. 99–116). Routledge.

Özturk, Y. (2018). *Discourse markers in native and non-native spoken English: A corpus-based comparison of Turkish and British university students' employment of discourse markers in informal interviews*, Unpublished PhD Thesis. Anadolu University.

Özturk, Y., & Köse, G. (2021). "Well (er) You Know …": Discourse markers in native and non-native spoken English. *Corpus Pragmatics*, *5*(2), 1–20. https://doi.org/10.1007/s41701-020-00095-9

Palmer, B.C., El-Ashry, F., Leclere, J.T. and Chang, S. (2007), Learning From Abdallah: A case study of an Arabic-speaking child in a U.S. school. *The Reading Teacher*, *61*(1), 8-17. https://doi.org/10.1598/RT.61.1.2

Pan, Z. & Aroonmanakun, W. (2022). A corpus-based study on the use of spoken discourse markers by Thai EFL learners. *LEARN Journal: Language Education and Acquisition Research Network, 15*(2), 187–213. https://eric.ed.gov/?id=EJ1358661

Paquot, M., & Tracy-Ventura, N. (2020). Second language acquisition and corpora. In M. Paquot & N. Tracy-Ventura (Eds), *The Routledge handbook of L2 acquisition and corpora* (pp. 1–7). Routledge.

Paradis, J. (2007). Second language acquisition in childhood. In E. Hoff & M. Shatz (Eds), *Blackwell handbook of language development*. Blackwell.

Paradis, J. (2011). Individual differences in child English L2 acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*, 213–237. https://doi.org/10.1075/lab.1.3.01par

Paradis, J. (2019). English L2 acquisition from early childhood to adulthood: The role of age, first language, cognitive, and input factors. In M. M. Brown & B. Dailey (Eds), *Proceedings of the 43rd Annual Boston University Conference on Language Development. Volume 2* (pp. 11–26). Conference on Language Development, Somerville, MA. Cascadilla Press.

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, *50*(4), 793–817. https://doi.org/10.1017/S0305000922000708

Paradis, J., Rusk, B., Duncan, T. S., & Govindarajan, K. (2017). Children's L2 acquisition of English complex syntax: the role of age, input, and cognitive factors. *Annual Review of Applied Linguistics*, *37*, 148–167. https://doi.org/10.1017/S0267190517000022

Passot, F. (2007). A because B so A: Circularity and discourse progression in conversational English. In A. Celle & R. Huart (Eds), *Connectives as discourse landmarks* (pp. 117–134). John Benjamins.

Philp, J., & Duchesne, S. (2008). When the gate opens: The interaction between social and linguistic goals in child L2 development. In J. Philp, R. Oliver, & A. Mackey (Eds), *Second language acquisition and the younger learner: Child's play?* (pp. 83–104). John Benjamins.

Philp, J. Mackey, A., & Oliver, R. (2008). Child's play? Second language acquisition and the younger learner in context. In J. Philp, R. Oliver, & A. Mackey (Eds), *Second language acquisition and the younger learner: Child's play?* (pp. 3-23). John Benjamins.

Polat, B. (2011). Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics*, *43*(15), 3745-3756. https://doi.org/10.1016/j.pragma.2011.09.009

Pons Bordería, S. P. (2008). Do discourse markers exist? On the treatment of discourse markers in Relevance Theory. *Journal of Pragmatics*, *40*(8), 1411–1434. https://doi.org/10.1016/j.pragma.2008.03.013

Ranger, G. (2018). *Discourse markers: an Enunciative approach*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-70905-5

Raymond, G. (2004). Prompting action: the standalone 'so' in ordinary conversation. *Research on Language and Social Interaction*, *37*(2), 185–218. https://doi.org/10.1207/s15327973rlsi3702_4

Rayson, P. (2015). Computational tools and methods for corpus compilation and analysis. In D. Biber & R. Reppen (Eds), *The Cambridge handbook of English corpus linguistics* (pp. 32-49). Cambridge University Press.

Rayson, P. (2019). Corpus analysis of key words. In C. Chapelle (Ed), *The concise encyclopaedia of applied linguistics* (pp. 320–326). Wiley.

Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, *14*(3) 367-381. https://doi.org/10.1016/0378-2166(90)90095-U

Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics*, *29*(6), 1139–1172. https://doi.org/10.1515/ling.1991.29.6.1139

Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. In K. Fischer (Ed), *Approaches to discourse particles* (pp. 339-358). Elsevier.

Ren, W. (2018). Developing pragmatic competence in study abroad contexts. In C. Sanz & A. Morales-Front (Eds), *The Routledge handbook of study abroad research and practice* (pp. 119–133). Routledge.

Rendle-Short, J. (2000). When “okay” is okay in computer science seminar talk. In *Australian Review of Applied Linguistics* (22) 2, 19–33. https://doi.org/10.1075/aral.22.2.02ren

Reppen, R., & Simpson-Vlach, R. (2020). Corpus linguistics. In N. Schmitt & M. P. H. Rodgers (Eds), *An introduction to applied linguistics* (pp. 91–108). Routledge.

Reyes, I. (2004). Functions of code switching in schoolchildren’s conversations. *Bilingual Research Journal*, *28*(1), 77–98. https://doi.org/10.1080/15235882.2004.10162613

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse Processes*, *14*(4), 423–441. https://doi.org/10.1080/01638539109544795

Romaine, S. & Lange, D. (1991). The use of *like* as a marker of reported speech and thought: A case of grammaticalisation in progress. *American Speech*, *66*(3), 227–279. https://doi.org/10.2307/455799

Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics*, *34*, 769–784. https://doi.org/10.1016/S0378-2166(02)00022-X

Romero-Trillo, J. (2018). Prosodic modelling and position analysis of pragmatic markers in English conversation. *Corpus Linguistics and Linguistic Theory*, *14*(1), 169–195. https://doi.org/doi:10.1515/cllt-2014-0026

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735. https://doi.org/10.2307/412243

Saville-Troike, M. (2012). *Introducing L2 acquisition*. Cambridge University Press.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of ‘uh huh’ and other things that come between sentences. In D. Tannen (Ed) *Analysing discourse: Text and talk, Georgetown University Roundtable on Languages and Linguistics* (pp. 71–93). Georgetown University Press.

Schegloff, E. A., Jefferson, G. & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382. https://doi.org/10.2307/413107

Schegloff, E. A., & Lerner, G. H. (2009). Beginning to respond: well-prefaced responses to wh-questions. *Research on Language and Social Interaction*, *42*(2), 91–115. https://doi.org/10.1080/08351810902864511

Schiffrin, D. (1985). Multiple constraints on discourse options: A quantitative analysis of causal sequences. *Discourse Processes*, *8*(3), 281–303. https://doi.org/10.1080/01638538509544618

Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press.

Schiffrin, D. (2001). Discourse markers: language, meaning, and context. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds), *The handbook of discourse analysis* (pp. 54–75). https://doi.org/10.1002/9780470753460.ch4

Schiffrin, D. (2006). Discourse marker research and theory: Revisiting *and*. In K. Fischer (Ed), *Approaches to discourse particles* (pp. 315–338). Elsevier.

Schleef, E. (2005). Gender, power, discipline, and context: On the sociolinguistic variation of *okay*, *right*, *like*, and *you know* in English academic discourse. In C. Sunakawa, T. Ikeda, S. Finch, & M. Shetty (Eds.), Texas linguistic forum (pp. 177-186). University of Texas.

Schourup, L. (1985). *Common discourse particles in English conversation*. Routledge.

Schourup, L. (1999). Discourse markers, *Lingua. 107*(3–4), 227-265. https://doi.org/10.1016/S0024-3841(96)90026-1

Schourup, L. (2001). Rethinking *well*. *Journal of Pragmatics*, *33*(7), 1025–1060. https://doi.org/10.1016/S0378-2166(00)00053-9

Schweinberger, M. (2014). *The discourse marker LIKE: a corpus-based analysis of selected varieties of English*. Unpublished PhD Thesis, University of Hamburg.

Schweinberger, M. (2023). On the L1-acquisition of the pragmatics of discourse *like*. *Functions of Language*, *30*(3), 255–286. https://doi.org/10.1075/fol.20025.sch

Scott, M. (1996). *Wordsmith Tools* (Computer software). Oxford University Press.

Scott, M. (1997). PC analysis of key words- and key key words. System, 25(2), 233–245. https://doi.org/10.1016/S0346-251X(97)00011-0

Selinker, L. (1972). Interlanguage. *International review of applied linguistics in language teaching*, *10*(3), 209-232. https://doi.org/10.1515/iral.1972.10.1-4.209

Serratrice, L. (2019). Becoming bilingual in early childhood. In L. Ortega & A. De Houwer (Eds), *The Cambridge handbook of bilingualism* (pp. 15–35). Cambridge University Press.

Siegel, M. (2002). *Like:* The discourse particle and semantics. *Journal of Semantics*, *19*(1), 35–71. https://doi.org/10.1093/jos/19.1.35

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Sinclair, J. & Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils; based on a report submitted to the Social Science Research Council*. Oxford University Press.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton.

Slabakova, R. (2016). *Second language acquisition*. Oxford University Press.

Smith, Y. (2019).Using corpus methods to investigate classroom interaction and teacher discourse in special educational needs (SEN) classrooms: an investigation of methodological possibilities, Unpublished PhD Thesis. Lancaster University.

Sopata, A. & Długosz, K. (2022). Age of onset effects in child bilingual acquisition: Identifying the turning point. *Gema Online Journal of Language Studies*, *22*(3), 1–23. https://doi.org/10.17576/gema-2022-2203-01

Sorenson Duncan, T., & Paradis, J. (2020). How does maternal education influence the linguistic environment supporting bilingual language development in child L2 learners of English? *International Journal of Bilingualism*, *24*(1), 46-61. https://doi.org/10.1177/1367006918768366

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and cognition*. Blackwell.

Sprott, R. (1992). Children's use of discourse markers in disputes: form-function relations and discourse in child language. *Discourse Processes*, *15*, 423–439. https://doi.org/10.1080/01638539209544821

Stefanowitsch, A. & Gries, S. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209–243. https://doi.org/10.1075/ijcl.8.2.03ste

Stenström, A. B. (1990). Lexical items peculiar to spoken discourse. In J. Svartvik (Ed), *The London-Lund Corpus of Spoken English: Description and Research* (pp. 137-157). Lund University Press.

Svartvik, J. (1980). *Well* in conversation. In S. Greenbaum, G. Leech, & J. Svartvik (Eds), *Studies in English linguistics: For Randolph Quirk* (pp. 167–177). Longman.

Svartvik, J. (1992). Corpus linguistics comes of age. In J. Svartvik (Ed), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991* (pp. 7–16). Nobel Symposium, Berlin ; New York. Mouton de Gruyter.

Tagliamonte, S. (2016). *Teen talk: The language of adolescents*. Cambridge University Press.

Tagliamonte, S. & Hudson, R. (1999). *Be like* et al. Beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics*, *3*(2), 147–172. https://doi.org/10.1111/1467-9481.00070

Taguchi, N. (2019). Second language acquisition and pragmatics: An overview. In N. Taguchi (Ed), *The Routledge handbook of L2 acquisition and pragmatics* (pp. 1-15). Routledge.

Tarone, E. (2018). Interlanguage. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-7). Wiley-Blackwell.

Taguchi, N. & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.

Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language learning*, *44*(2), 307–336. https://doi.org/10.1111/j.1467-1770.1994.tb01104.x

Thuy, D. (2019). A Relevance-Theoretic analysis of *yeah* as a discourse marker. *VNU Journal of Foreign Studies*, *35*, 176-186. https://doi.org/10.25073/2525-2445/vnufs.4377

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins.

Tognini-Bonelli, E. (2012). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds), *The Routledge handbook of corpus linguistics* (pp. 14-27). Routledge.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University press.

Traugott, E. C. (1982). From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalisation. In W. P. Lehmann & Y. Malkiel (Eds) *Perspectives on Historical Linguistics* (pp. 245–271). John Benjamins.

Traugott, C. (1995). T*he Role of the Development of Discourse Makers in a Theory of Grammaticalisation.* Resource Document, Paper Presented at ICHL XII. http://www.wata.cc/forums/uploaded/136_1165014660.pdf

Underhill, R. (1988). Like is, like, Focus. *American Speech*, *63*(3), 234–246. https://doi.org/10.2307/454820

Unsworth, S. (2008). Comparing child L2 development with adult L2 development: How to measure L2 proficiency. In B. Haznedar & E. Gavruseva (Eds) *Current trends in child L2 acquisition: A generative perspective* (pp. 301–333). John Benjamins. https://doi.org/10.1075/lald.46.15uns

Unsworth, S. (2016a). Early child L2 acquisition: Age or input effects? Neither, or both? *Journal of Child Language*, *43*(3), 608–634. https://doi.org/10.1017/S030500091500080X

Unsworth, S. (2016b). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 136-196). Mouton De Gruyter.

Van Dijk, T. A. (1979). Pragmatic connectives. *Journal of Pragmatics*, *3*(5), 447–456. https://doi.org/10.1016/0378-2166(79)90019-5

Van Olmen, D., & Šinkuniene, J. (2021). Introduction: Pragmatic markers and peripheries: An overview. In D. Van Olmen & J. Šinkuniene (Eds), *Pragmatic markers and peripheries* (pp. 1–16). John Benjamins.

VanPatten, B., & Willams, J. (2015). Early theories in SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 17–33). Routledge.

Watts, R. J. (1989). Taking the pitcher to the 'well': Native speakers' perception of their use of discourse markers in conversation. *Journal of Pragmatics*, *13*(2), 203–237. https://doi.org/10.1016/0378-2166(89)90092-1

Weirong, C. (2017). Functions of so in English as a lingua franca and native speech: A corpus-based study. *Chinese Journal of Applied Linguistics*, *40*(1), 93–111. https://doi.org/doi:10.1515/cjal-2017-0006

Wilson, D. (2011). The conceptual-procedural distinction: Past, present and future. In V. Escandell-Vidal, M. Leonetti, & A. Ahern (Eds), *Procedural meaning: Problems and perspectives* (pp. 3–31). Emerald Group.

Wilson, D., & Sperber, D. (2002). Relevance Theory. In L. Horn & G. Ward (Eds), *The Handbook of Pragmatics* (pp. 607–632). Blackwell.

Wong, J. (2000). The token 'yeah' in non-native speaker English conversation. *Research on Language and Social Interaction*, *33*(1), 39–67. https://doi.org/10.1207/S15327973RLSI3301_2

Xiao, R. (2015). Collocations. In D. Biber & R. Reppen (Eds), *The Cambridge handbook of English corpus linguistics* (pp. 106-124). Cambridge University Press.

Yeh, K., & Huang, C. (2016). Mandarin-speaking children's use of the discourse markers *hao* 'okay' and *dui* 'right' in peer interaction. *Language Sciences*, *57(1)*, 1–20. https://doi.org/10.1016/j.langsci.2016.04.004

Zahler, S. (2023). Some issues in usage-based methods: Contributions from corpus linguistics, psycholinguistics, and variationist sociolinguistics. In M. Díaz-Campos & S. Balasch (Eds.), *The handbook of usage-based linguistics* (pp. 73–90). John Wiley & Sons, Inc.

Zwicky, A. M. (1985). Clitics and particles. *Language*, *61*(2), 283–305. https://doi.org/10.2307/414146

Appendix A: PARTICIPATION FLYER & INVITATION NOTICE

Lancaster University

Dear parent/ guardian(s)

I would like to invite you and your child to participate in a research study as part of my PhD project. The research aims at studying the acquisition of L2 through the use of linguistic items known as discourse markers.

Your child participation will involve being a part of a free conversation with one of her classmates for a period not more than ten minutes. The free conversation will be audio recorded only. We also need some information about your child background which we will ask you kindly to fill in a form of questionnaire later.

You and your child are not obliged to take part in the study. If you and your child agree to participate, you will be free to withdraw from the study at any time without any obligation to explain your reasons.

If you and your child decide to take part of the study or not, please write your child name and grade and send this letter back with your child to her school teacher.

Please do not hesitate to contact me if you would like to discuss the information provided or ask any questions before agreeing to be part of the study.

Yours sincerely

Nuha Alharbi

PhD student at Lancaster University

Email: n.alharbi @lancaster.ac.uk

☐ Yes, I and my child would like to participate in the study.

☐ No, we are not interested to be part of the study.

Your child name: ……

Grade: …….

عزيزي ولي الأمر / الوصي (الأوصياء)

أرغب بدعوتك ودعوة ابنتك للمشاركة في دراسة بحثية كجزء من مشروع الدكتوراه الخاص بي. يهدف البحث إلى دراسة اكتساب لغة ثانية من خلال استخدام مفردات لغوية تعرف بعلامات الخطاب.

ستشمل مشاركة طفلك إجراء محادثة عفوية مع أحدى زميلاتها في الفصل لمدة لا تزيد عن عشر دقائق. سيتم تسجيل المحادثة المجانية بالصوت فقط. نحتاج أيضًا إلى بعض المعلومات الشخصية عن ابنتكم ، وسنطلب منكم التفضل بملء نموذج استبيان لاحقًا.

علماً أنك وابنتك غير ملزمين بالمشاركة في الدراسة. إذا وافقتم على المشاركة ، فسيكون لكم الحرية في الانسحاب من الدراسة في أي وقت دون أي التزام بشرح أسباب الانسحاب.

إذا قررت أنت وطفلك المشاركة في الدراسة أم لا ، فيرجى كتابة اسم طفلك وفصلة وإرسال هذه الرسالة مع ابنتكم إلى المدرسة.

الرجاء عدم التردد في الاتصال بي إذا كنت ترغب في مناقشة أي من المعلومات المقدمة أو طرح أي أسئلة قبل الموافقة على أن تكون جزءًا من الدراسة.

تفضلوا بقبول فائق الاحترام

نهى الحربي

طالبة دكتوراه في جامعة لانكستر

بريد إلكتروني: n.alharbi @ lancaster.ac.uk

☐ نعم ، أنا وطفلي نرغب في المشاركة في الدراسة.

☐ لا ، لسنا مهتمين بأن نكون جزءًا من الدراسة

اسم الطالبه: .....

الفصل:.....

Children at grade (four, five and six) are needed for a research study

Want to Participate

If you are a girl and aged between 9- and 12-year-old, we would like to invite you to participate in our study. All you are going to do is to have a chat with your classmate and we will record that.

You are going to have this conversation during the school time.

No preparation is required

Appendix B: INFORMATION SHEET

Linguistics & English Language | Lancaster University

Participant information sheet

I am Nuha Alharbi, a lecturer at Prince Sattam bin Abdulaziz University in Al-Kharj, Riyadh and a PhD student at Lancaster University. I would like to invite you to take part in a research study about "Corpus-based study of discourse markers in L2 English by L1 Arabic children".

Please take time to read the following information carefully before you decide whether or not you wish to take part.

What is the study about?

The study focus on how Arabic children are acquiring English as a L2 through their use of linguistic features known as discourse markers in their daily communication.

Why have I been invited?

As you are one of the parents whose children are studying in a Saudi international school, I would like to invite you to give consent for your child to participate in the study.

I have approached you because I am trying to understand how Saudi international school children are using English as a L2 in their conversations.

I would be very grateful if you would agree to take part in this study.

What will I be asked to do if I take part?

If you decided to take part, this would involve the following:

- On behalf of your child, you will fill in a background questionnaire about your child. Your child will participate in a ten minute conversation with one of their classmates during the school day.

Participation in the study will not disrupt any other part of your child's school schedule.

What are the possible benefits from taking part?

Your child's participation will contribute to our understanding of how Arabic children acquire a L2 in an intensive L2 environment. This will also help teachers and teaching practitioners in designing teaching materials to encourage children to become more fluent in their L2.

Does my child have to take part?

No. It's completely up to you and your child to decide whether or not they take part. Participation is voluntary and your child is free to withdraw at any time, without giving any reason. Additionally, if you decide not to let your child take part in this study, this will not affect their studies and the way they are assessed on any course.

What if I change my mind?

If you (or your child) take the decision to withdraw, I will extract any data they contributed to the study and destroy it. Data means the information, views, ideas, etc. that your child and other participants will have shared with me.

What are the possible disadvantages and risks of taking part?

There are no anticipated risks in taking part.

Will my data be identifiable?

After recording children's conversations, only I, as the researcher conducting this study, and my research supervisors will have access to the data they share with me.

I will keep all personal information about your child (e.g. their name and any other information that can identify them) confidential, that is I will not share it with others. I will anonymise any audio recordings and hard copies of any data. This means that I will remove any personal information.

How will my data be stored?

The data will be stored on a secure server at Lancaster University. Access to the data will be restricted to me and my research supervisors. Data is kept securely for a period of ten years, in accordance with University guidelines.

How will we use the information you have shared with us and what will happen to the results of the research study?

I will use the data for academic purposes only. This means that it will be included in my reporting of the study, which will include my PhD thesis and can include publications such as journal articles, as well as presentations at academic conferences.

When writing up the findings from this study, I will anonymise any quotations from the conversations involving your child, in order to protect their identity.

Who has reviewed the project?

This study has been reviewed and approved by the Faculty of Science and Technology Research Ethics Committee.

What if I have a question or concern?

If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact myself or my supervisors at the details below.

Nuha Al-harbi **PhD student**	**Dr Luke Collins** Senior Research Associate	Dr Andrew Hardie Reader in Linguistics
Department of Linguistics and English Language Lancaster University	Department of Linguistics and English Language Lancaster University	Department of Linguistics and English Language Lancaster University
E-mail: n.alharbi@lancaster.ac.uk	E-mail: l.collins3@lancaster.ac.uk	E-mail: a.hardie@lancaster.ac.uk

If you have any concerns or complaints that you wish to discuss with a person who is not directly involved in the research, you can also contact:

Professor Jonathan Culpeper
Head of Department
Department: Linguistics and English Language
E-mail: j.culpeper@lancaster.ac.uk
Phone number: *+44 (0)1524 593045*
County South
Lancaster University
Lancaster
United Kingdom
LA1 4YL

Thank you for considering your participation in this project.

Appendix C: BACKGROUND QUESTIONNAIRE

Linguistics & English Language | Lancaster University

Child Background Questionnaire

Please, fill in the sheet with the required information. You can choose to fill out the form in English OR in Arabic (see reverse). If you require any further information, please contact the researcher Nuha Alharbi n.alharbi@lancaster.ac.uk

Thank you for participating in the study.

Your child's first name:
Your child's Surname:
School name:
Age:
Nationality:
Number of siblings:
Main Spoken language at home: Arabic ☐ English ☐ Other Please specify _____
Mother's native language_____
Father's native language_____
Did your child attend any preschool or nursery? Yes ☐ No ☐ If yes, was their study in Arabic or English: _____
Did your child spend any time in countries speaking English as a native language? Yes ☐ No ☐ If yes, how much time did they spend there? _____ Did they join any formal education there? If yes, please specify all of them. _____ _____

Linguistics & English Language | Lancaster University

معلومات عن المشاركة

أرجو منكم التكرم بملء الخانات بالمعلومات المطلوبة

في حال واجهتكم بعض الصعوبات في ملء البيانات بإمكانكم التواصل مع الباحثة

أشكر لكم مشاركتكم في هذه الدراسة

الباحثة: نهى الحربي

n.alharbi@lancaster.ac.uk

الاسم الاول لإبنتكم:
اللقب:
المدرسة:
عمر الابنه:
الجنسية:
عدد الأخوة:
اللغة الأساسية في المنزل العربية ☐ الأنجليزية ☐ ☐ أخرى أذكرها ---------- ☐
اللغة الاساسية للأم ----------
اللغة الأساسية للأب----------
هل التحقت الابنه بمرحلة رياض الأطفال نعم ☐ لا ☐
إذا كانت الأجابة بنعم فهل تمت الدراسة باللغة العربية أم الإنجليزية ----------
هل قضت ابنتكم أي وقت في دولة تتحدث اللغة الأنجليزية كلغة أساسية نعم ☐ لا ☐
إذا كانت الإجابة بنعم فكم المدة التي قضتها هناك وهل التحقت بإي مرحلة دراسية هناك، إذا كانت الإجابة نعم – أرجو تحديد المراحل الدراسية التي تمت دراستها هناك ----------

Appendix D: PROMPT CARDS

Family	Holiday	Sport
Movies	Shopping	Stories (funny, scary, sad)
School subjects	food	Animals
Travel	Songs	Games
Books	Pets	Friends

Appendix E: AR-CHILDREN TRANSCRIPTION CONVENTIONS AND XML TAGS

Feature	Description	Transcription	XML
Speaker ID	Name of speaker with the time and date of recording	Speaker 1	<u who= “080201”>
Anonymised name	Name of participants, family members, friends or teachers	<name>	<anon type="name">
Anonymised place	places where participants live or lived include neighbourhood, cities or countries	<place>	<anon type="place">
Anonymised birth date			<anon type=" dateOfBirth ">
Short pause	Pause less than 5 seconds	<short pause>	<pause dur="short"/>
Long pause	Pause equal 5 seconds or more.	<pause>	<pause dur="long"/>
Laughter		<laugh>	<vocal desc="laugh" />
Cough		<cough>	<vocal desc="cough" />
Overlap	Overlaps are not marked		
Filled pauses	Only five fillers are included	*mm, uhu, uh, ah,* and *oh*	

The tags in the Ar-Children corpus, taken from the BNC2014 manual (Love et al., 2017)

Appendix F:DMs ACCORDING TO EACH INDIVIDUAL SPEAKER'S USE AND THE LENGTH OF TIME SPENT IN PRESCHOOL

main IDs	No. of words	*like*	*so*	*well*	*yeah*	*okay*	*oh*	*you know*	*I mean*	age	Time abroad
080201	1102	17	0	0	12	2	2	0	0	11	0
080202	499	15	3	0	2	11	6	0	1	11	2
082101	846	32	4	0	7	1	3	1	0	11	1
082102	287	2	1	0	6	0	0	0	0	12	0
082401	863	12	1	1	6	11	4	14	0	10	2
082402	579	12	0	0	2	1	2	0	0	10	0
083301	1247	10	11	0	2	5	2	0	0	10	2
083302	398	5	3	0	1	3	0	0	0	10	1
083401	367	1	1	0	4	4	1	0	0	9	0
083402	857	14	7	0	6	5	7	1	0	9	0
084601	589	6	8	0	0	12	2	1	0	9	2
084602	420	0	11	1	2	5	3	0	0	9	0
084801	1552	26	5	4	14	7	1	4	1	11	1
084802	1258	28	3	1	11	5	5	2	1	11	1
085301	729	1	2	2	5	0	3	0	1	10	2
085302	293	2	1	0	4	1	0	0	0	9	0
085601	602	1	9	0	0	3	6	0	0	9	0
085602	491	14	3	0	0	0	0	0	0	9	0
090201	581	18	6	0	12	15	14	8	0	10	0
090202	1415	24	8	1	8	3	3	5	0	10	3
090601	969	50	5	1	7	3	0	0	0	9	2
090602	586	3	8	0	19	0	4	0	0	9	1
090801	335	1	1	0	9	0	1	0	0	9	0
090802	481	5	9	0	3	1	0	0	1	9	0
091601	500	0	1	1	3	11	2	0	0	10	0
091602	621	5	3	0	4	3	1	0	0	11	0
091701	774	11	6	0	10	2	0	0	1	9	2
091702	431	7	2	1	5	8	3	0	0	9	0
091801	328	1	1	0	7	3	0	0	0	11	0
091802	509	5	4	1	3	0	0	0	1	12	0
092901	479	4	3	2	9	6	3	0	0	11	2
092902	586	8	2	0	10	2	8	0	0	11	1
093001	581	2	3	0	3	2	1	0	0	10	0
093002	646	2	5	0	6	10	6	0	0	10	1
094101	1038	30	10	0	5	11	7	3	0	9	1
094102	652	5	1	1	9	4	3	0	0	9	0
094501	722	10	2	3	6	2	3	3	0	10	0
094502	670	0	6	0	8	12	1	0	0	10	0

094601	842	23	1	0	20	0	0	1	1	11	0
094602	983	8	1	2	25	2	19	1	0	11	2
095801	1002	0	0	0	2	2	2	1	0	10	0
095802	666	10	0	1	1	4	3	0	0	11	0
101001	426	5	0	0	0	8	0	0	0	11	0
101002	322	0	0	0	0	0	0	0	0	12	1
101201	628	6	12	0	1	10	1	0	0	9	0
101202	872	17	12	0	3	11	2	0	0	9	0
101501	532	9	3	0	6	9	3	0	0	12	0
101502	309	0	0	0	1	2	0	0	0	11	0
101801	856	5	7	6	9	0	7	1	1	11	3
101802	1036	9	5	5	7	3	18	0	5	12	2
102301	948	32	4	0	3	1	0	1	1	10	2
102302	369	5	0	0	3	9	7	0	0	10	1
102601	754	12	0	0	2	5	1	0	0	11	0
102602	729	10	7	0	0	8	1	0	0	12	0
103701	650	2	3	1	0	0	2	4	0	11	1
103702	972	24	6	0	20	0	5	2	0	12	0
103801	790	11	4	0	0	2	1	0	1	12	0
103802	415	1	0	0	6	8	1	0	0	11	0
104801	695	0	1	0	0	1	0	0	0	11	0
104802	602	3	0	0	3	14	5	0	0	12	0
110101	525	5	2	0	8	3	1	0	0	9	1
110102	584	3	0	0	3	3	2	0	0	9	0
110201	818	10	0	1	8	7	2	0	0	12	2
110202	722	14	0	1	10	0	0	6	0	11	1
110501	272	2	1	0	0	2	0	0	0	10	2
110502	509	1	2	0	1	5	0	0	1	10	0
110701	852	11	9	0	4	5	1	0	0	9	1
110702	372	1	1	0	13	7	1	0	0	9	0
111601	296	1	2	0	0	0	2	0	0	10	0
111602	955	14	7	0	7	10	5	1	0	10	2
111901	435	2	0	1	10	0	0	0	0	9	1
111902	760	11	1	0	13	4	0	2	0	9	0
112101	439	1	1	0	0	11	0	0	1	9	0
112102	259	0	0	0	2	0	0	0	0	9	0
112301	534	0	1	0	0	4	0	0	0	11	0
112302	466	0	0	0	0	0	1	1	0	10	0
112701	825	50	0	0	6	2	2	0	0	9	0
112702	453	3	0	0	4	0	2	0	1	9	2
113201	680	10	1	0	9	1	0	3	0	9	0
113202	421	6	0	0	9	1	0	1	0	9	1
113401	672	6	1	0	0	12	1	0	0	9	0

113402	100	0	0	0	0	1	0	0	0	9	0
113501	566	3	2	0	3	1	2	0	0	9	0
113502	788	4	0	0	8	1	0	1	0	9	2
114601	536	0	1	1	0	3	1	0	0	10	3
114602	300	0	0	1	2	2	1	0	0	9	0
114701	565	0	0	0	0	7	2	0	0	10	3
114702	200	0	0	0	0	0	0	0	0	10	0
115301	686	3	2	0	3	4	3	0	1	9	1
115302	311	0	3	0	1	4	0	0	0	9	0
115701	1403	14	2	0	1	0	0	0	1	9	2
115702	325	2	0	0	4	0	0	1	0	9	0
120801	1306	6	9	0	10	5	10	8	1	10	1
120802	717	35	0	1	8	1	3	1	0	10	1
121401	888	32	0	0	17	24	1	0	0	9	1
121402	560	19	1	0	7	17	2	0	0	9	0
122001	552	18	2	0	5	9	4	1	0	9	2
122002	743	4	4	0	2	2	2	1	0	10	0
122401	560	1	0	0	1	8	1	0	2	11	0
122402	407	0	6	0	1	1	0	0	0	11	2
123001	447	0	1	0	1	3	0	1	0	10	0
123002	439	1	0	0	0	3	0	0	0	10	1
123501	518	10	0	0	0	8	0	0	0	11	0
123502	548	8	1	0	4	5	0	0	0	11	1
124301	465	11	0	0	2	1	0	0	1	10	0
124302	223	0	0	0	0	1	0	0	0	10	0
124501	500	21	2	0	2	25	0	0	0	12	0
124502	554	13	5	0	12	22	2	0	0	11	0
125701	922	17	4	0	15	1	2	1	0	12	1
125702	623	29	3	0	21	1	11	1	0	11	0
132001	374	9	0	0	0	3	0	0	0	11	0
132002	444	14	0	0	2	10	2	0	0	11	0
392901	947	21	2	0	3	11	6	0	0	9	1
392902	1172	29	4	1	6	0	1	0	0	9	1
693001	1036	58	2	2	7	1	13	0	0	11	1
693002	1163	37	3	0	14	0	13	2	0	10	2
835001	1189	19	5	1	4	7	1	1	0	10	0
835002	275	4	0	0	1	1	0	0	0	10	0
958601	1002	16	2	1	15	1	0	0	0	11	2
958602	551	9	1	0	8	4	6	0	0	11	0

Appendix G: FREQUENCY LISTS OF TOP 50 FREQUENT WORDS IN AR-CHILDREN & BNC2014 CHILD SUBCORPUS

	Ar-Children		BNC2014 children subcorpus	
No.	**Word**	**Frequency**	**Word**	**Frequency**
1	,	9,433	i	6,033
2	/	4,375	it	3,831
3	.	4,222	's	3,311
4	i	3,252	and	3,281
5	and	2,103	the	3,182
6	like	2,069	you	3,022
7	the	1,676	a	2,884
8	you	1,660	yeah	2,653
9	to	1,299	?	2,590
10	it	1,269	that	2,352
11	uh	1,266	like	2,306
12	's	1,107	to	2,104
13	my	1,097	--unclearword	1,870
14	do	1,042	n't	1,794
15	mm	929	do	1,598
16	what	924	they	1,354
17	yeah	848	he	1,330
18	a	815	no	1,311
19	in	752	of	1,301
20	is	750	was	1,265
21	so	642	is	1,147
22	was	620	in	1,122
23	your	606	mm	1,074
24	n't	590	oh	1,068
25	that	566	but	1,067
26	okay	560	just	1,033
27	favourite	503	have	1,004
28	me	496	so	952
29	love	495	erm	922
30	have	491	then	916
31	we	484	we	905
32	but	469	what	864
33	she	442	not	837
34	know	426	there	835
35	go	410	well	805

36	no	380	know	747
37	of	348	er	743
38	then	341	be	728
39	oh	317	can	718
40	or	313	'm	710
41	one	306	on	709
42	with	304	really	674
43	when	290	--anonnamem	672
44	they	288	're	670
45	her	274	my	670
46	because	269	think	668
47	there	254	one	660
48	this	250	if	613
49	did	242	this	594
50	not	225	cos	563