University of Alberta

Constructions in child second language acquisition:
Exploring the role of first language and usage

by

Tatiana Zdorenko

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

Spring 2011
Edmonton, Alberta

Abstract

This thesis examined the factors of L1, input frequency and emergent productivity in child L2 acquisition. These issues were addressed in many previous studies on L1 acquisition and adult L2 acquisition, but this thesis is the first study to look at the interplay of L1 and usage factors in children learning a L2. The focus of the thesis was an investigation of these factors in the acquisition of article and auxiliary systems of English, which have been proven to be problematic areas for both L1 and L2 learners. While accounts of L1 transfer in L2 acquisition are better developed in generative theory, the roles of input frequency and emergent productivity are better developed in constructivist theory. The thesis assessed these two approaches against the data from L2 children from various L1 backgrounds. The children's accuracy and error patterns with articles and auxiliaries were investigated over two years of development.

The main findings were as follows. L1 typology facilitated the acquisition of the structure of the Noun Phrase (NP) and the Verb Phrase (VP), but it only extended as far as the awareness of the presence of the functional morpheme (article or auxiliary). The particular structure of the NP and VP in L1s of the children did not have a significant effect on their performance. L1 transfer effects were observed only in the first 1.5 years of acquisition. It was argued that the short-lived nature of L1 transfer could be due to the unstable L1 knowledge in child L2 learners. The use of articles and auxiliaries was also influenced by their input frequencies and distribution, as more frequent forms were supplied more

accurately and were substituted for less frequent forms. Different forms of articles and auxiliaries emerged separately and followed different paths of development. It was argued that they were acquired piecemeal as parts of constructions and that productivity with these forms emerged gradually.

It was concluded that constructionist theories were better supported by the data, since the findings on input frequency and productivity were not compatible with the generative approach, and L1 transfer was incorporated into the constructionist approach to account for the findings. It was argued that by the onset of L2 acquisition, child L2 learners had established constructions in their L1 that were abstract enough to be transferred to L2 and did not rely on lexically specific information. All children learned morphological forms of L2 in piecemeal fashion, and doing so they demonstrated input effects that held across all L1 backgrounds.

Table of contents

CHAPTER 1. INTRODUCTION

1.1 Background: Defining child L2 acquisition

Research on child second language (L2) acquisition can offer insight into both first language (L1) acquisition and L2 acquisition. First of all, L2-acquiring children, while cognitively more mature than children learning their L1, are still likely to achieve native-like competence in L2, hence it is possible that the same mechanisms are operative in child L2 as in L1 acquisition, and studying child L2 can contribute to theoretical debates pertaining to L1. On the other hand, the L2 child has a considerable knowledge of another language at the outset of acquisition, so it possible to test to what extent child L2 acquisition is influenced by L1 transfer – a phenomenon that has received much attention in the field of L2 acquisition. Thus, the study of child L2 acquisition has much to offer the field of language acquisition.

The term child L2 acquisition will be used to refer to L2 learners who start the acquisition of a foreign language between the ages of 4 and 7. The rationale behind setting the age of the onset of acquisition to 4 years is simply to draw a line between simultaneous acquisition of two languages from birth (bilingualism) and the acquisition of an L2 after some knowledge of L1 grammar is already in place (Schwartz, 2003, 2004; Unsworth, 2005), although some studies have chosen 3 years as the lower age boundary (MacLaughlin, 1978; Lakshmanan, 1995). The cut-off point between first language (L1) and L2 acquisition still remains rather arbitrary, since there is a lack of studies that would empirically determine where the line should be drawn. The upper cut-off point of seven years was chosen following the conclusions in Johnson & Newport (1989) and DeKeyser (2000). These studies provided empirical evidence that learners who start the acquisition of a L2 before the age of 7 years are likely to become native speakers of this language. Thus, with this cut-off age we can expect less individual variation with respect to the learners' progress and no cases of

fossilization. It needs to be borne in mind, however, that the age of 7 years strictly speaking does not distinguish child L2 learners from adult L2 learners.

1.2 Major theoretical approaches

Knowledge of a language and the process of language acquisition are viewed quite differently by the researchers working in two opposing theoretical frameworks, the generativist framework and the emergentist framework. There are several theoretical models under the general umbrella of emergentism (O'Grady, 2008; O'Grady, Lee & Kwak, 2009), and the usage-based model is the one considered in this study.

The generativist and the emergentist position differ with respect to their views on how language is structured and the process of language acquisition. The generative position with respect to syntax is that there are innate syntactic categories, and the emergentist position is that there is no language-specific knowledge that children are equipped with from birth. In the generative framework, grammar is assumed to be a system of abstract rules that operate over categories of linguistic items. The knowledge of abstract linguistic categories is assumed to be a part of the universal grammar (UG), an innate blueprint that delimits possible grammars of natural languages. In the generative view, language acquisition is perceived as a task of mapping the words of the input language onto the pre-given abstract categories from the universal inventory. On both the generative and the emergentist position, the final state includes syntactic categories and features, such as noun, verb, etc. (the lexical categories), and tense, person, determiner, etc. (functional categories and features). The main difference between the two approaches concerns whether language learning operates over uncategorized words and word combinations and yields categories as on outcome or whether the goal of language learning is to match uncategorized words to categories that are innate. Emergentism takes the first position, that categories are constructed. Generativism takes the second position, that words are mapped onto categories. Emergentism argues that speakers of a language operate with a variety

of structures (constructions) that can involve not only abstract categories but also partially or fully lexically specific elements, and that structures of a language represent a collection of constructions ranging from fully abstract to fully lexically specified[1].

Emergentism is not one approach but rather a set of approaches that oppose the UG program. The approach within the emergentist framework that focuses on describing various patterns in the way language is structured is called constructionism (Croft, 2001; Goldberg, 1995; MacWhinney, 1999; Tomasello, 2003; Tomasello & Brooks, 1999). There is also emergentist work that focuses on the way language is acquired, and it can be divided into two types, depending on the focus of investigation. One is the usage-based approach, which focuses on the importance of the input (or usage) for understanding how language acquisition works. This approach has gathered a large body of research (Bybee, 1995; Ellis, 2002; 2006; Goldberg, 1995; Langacker, 1988). The other approach is a smaller body of research that explores the role of the processor-working memory interface in language acquisition (O'Grady, 2008). Most usage-based studies focused on L1 acquisition, but more recently, there has been growing interest in the relevance of the usage-based model to understanding L2 acquisition as well (Ellis & Larsen-Freeman, 2006, O'Grady et al., 2009). Thus, when considering child L2 acquisition from the point of view of emergentism, this thesis will focus in particular on the usage-based approach.

Usage-based approach applied to L1 acquisition

In the case of L1 acquisition, the proposal of the usage-based approach is that the child initially learns individual, lexically specific linguistic constructions (e.g. 'verb island constructions', Tomasello, 1992; 2000). Language learning, in this view, proceeds in a piecemeal fashion as the analysis of patterns, progressing from memorized chunks of one or two words to more complex phrases that have

[1] There is a variety of terms used in emergentism for various types of structures. Fully abstract constructions can also be called *schemas*, partially specified constructions are also called *item-based constructions*, and fully lexically specified constructions are called *chunks* or *formulas*. In this thesis, the term *construction* (*abstract* or *lexically specific*) will be used for all these structures.

slots into which the child can place a variety of words (e.g., *I don't* + Verb; *where's* + Noun?). As children become more and more productive with these patterns, they generalize across them, creating more abstract constructions. This general approach is usage-based in the sense that all linguistic knowledge – however abstract it may ultimately become – derives from the comprehension and production of specific utterances on specific occasions of use.

Importantly, both the generativist and the usage-based approach argue that there are some abstract categories in the final state of language acquisition. However, they view the notion of 'abstract category' differently. In generativist theories, UG is thought to equip children with knowledge of abstract properties such as agreement, tense, or case, and abstract operations such as 'move'. Language specific properties are learned by children from the input, but they are linked with the abstract innate knowledge from UG. In the emergentist view, grammar is seen as a collection of constructions, and the abstractness of a construction depends on how much variability is permitted in its elements. Abstractness of elements (or slots) in constructions is a matter of degree and higher variability results in higher productivity of a construction. Importantly, abstract constructions are not automatically assumed to be the "core'' of grammar, like the abstract categories and operations from a generative grammar point of view.

The theoretical difference in the view of what constitutes abstract knowledge has implications for language acquisition. The emergentist theory posits that in the beginning children make virtually no linguistic abstractions at all, only later attempting to zero in on adult-like linguistic categories and constructions. The generativist approach argues that learners cannot induce abstractions and that in order to recognize similarities among different lexical items a learner must already have the abstraction a priori, that is, innately.

Usage-based approach applied to L2 acquisition

All theoretical approaches acknowledge the fact that a L2 learner, child or adult, is different from a L1 learner in many respects. First, L2 learners are more

cognitively mature: In L1 acquisition, knowledge of the world and knowledge of language are developing simultaneously, whereas adult and child L2 acquisition builds on preexisting conceptual knowledge. Secondly, L2 learners already know one language, and naturally L2 acquisition is expected to build on the preexisting knowledge of a L1 as well. But because of crucial theoretical differences between the two frameworks, they give different value to the issues that influence the process of acquisition, both L1 and L2. The following section will briefly mention the issues that will be addressed in this thesis from the two theoretical perspectives, and Chapter 2 will provide a detailed discussion.

1.3 Issues to be addressed in child L2 acquisition

1. The role of L1, i.e. the extent to which children's acquisition reflects the structures of their L1. One of the most influential views of L1 transfer in the generative framework is the Full Transfer/Full Access model (Schwartz & Sprouse, 1996). This model hypothesizes that the initial state of L2 acquisition is the entire grammar of the learner's L1 (full transfer) and that in the process of learning the grammar gets restructured, never violating the universal principles (full access to UG). The restructuring of the initial L1 grammar happens when it fails to account for the L2 input that the L2 learner receives. Among the studies that provided evidence for this model were Haznedar (2001), Slabakova (2000), and Yuan (1998).

L2 acquisition in general and the issue of L1 transfer in particular have not been systematically addressed in usage-based studies. Ellis (2002: 170) mentioned that the process of moving from unanalyzed chunks to abstract generalizations is central in L2 acquisition, just like in L1 acquisition. In a usage-based perspective, there seems to be no reason to treat the developmental path of L1 and L2 acquisition differently: "The L1 acquisition sequence—from formulas, through low-scope patterns, to constructions—could serve well as a reasonable default in guiding the investigation of the ways in which exemplars and their type and token frequencies determine the second language acquisition of structure" (Ellis, 2002:

170). From the emergentist point of view, language (first or second) is acquired through usage, by extracting patterns and regularities from the input.

So far researchers working in the emergentist framework limited themselves to simply stating that L1 may play a role in L2 acquisition (Bybee & Tomasello, 2008; Ellis & Larsen-Freeman, 2006). It is not very clear then how an emergentist approach would accommodate the observation that, for instance, Spanish-speaking learners of L2 English have an advantage over Japanese-speaking learners because Spanish is structurally more similar to English than Japanese. If L1 knowledge is perceived as a network of constructions of varying degrees of entrenchment, how can this knowledge influence the morphosyntax of the L2? L1 influence on L2 structures needs to be investigated further in this perspective.

2. The role of speech that children hear (input). The usage-based approach expects a close relation between input and the patterns in acquisition. In the generative perspective, the role played by the input is the one of lexical learning that must take place in order for language-specific words to be mapped onto abstract categories. To illustrate the difference, if we take auxiliary BE, the usage-based approach would predict that the more frequent form *is* will be used sooner than *are*, there will be more errors in the contexts for *are* than for *is*, and the more frequent form will be often used as the substitute for the less frequent form. But in the case of the generative approach, no protracted differences between two forms of the same auxiliary are expected because the abstract categories such as Tense and Agreement are present from the onset. Once the two forms are known (from the input), they should be placed in the structure correctly.

3. The pattern of emergent productivity. The process of increasing productivity in language acquisition is at the heart of the usage-based framework. This process is viewed as piecemeal learning of constructions (words or word combinations) resulting in gradual emergence of abstract knowledge, and it proceeds as follows. Children associate forms early on only with the contexts in which they have heard

them. From those contexts, they must abstract out general patterns applicable across multiple contexts in order to use forms in novel situations. Similarly, children learn isolated linguistic structures as rote-learned wholes that they gradually learn to chunk together, to break into smaller component parts, and to relate to each other. For example, a child might first use different forms of go (*gone*, *went*, *going*) in distinct constructions before learning they are all related.

In contrast, the generativist view is a words-and-rules approach to learning. Individual words need to be learned and immediately linked to abstract categories that are already known (Pinker, 1999). Generativists do acknowledge that certain memorized formulas and idioms are learned as wholes, but these are considered exceptional and equated to learning of individual words.

In sum, the three issues discussed in this section have not been equally developed in the two theoretical frameworks. An approach to L1 transfer is relatively better developed in generative theories, but the role of input and the growth in productivity are relatively better developed in emergentist theories. The primary focus of the thesis is to explore these three issues from the two theoretical points of view using the data from child L2 acquisition of articles and auxiliaries, and in so doing, contribute to both empirical and theoretical aspects of language acquisition. While L1 transfer will be analyzed in the well-established generative models and will provide some support for them, in the end, suggestions will be put forward about how L1 transfer can be analyzed in an emergentist model. This will be done following the conclusions that article and auxiliary acquisition overall provides more support for the emergentist models and thus, L1 transfer needs to be added into these models to provide a comprehensive description of child L2 acquisition.

1.4 Objectives for this thesis

General theoretical aims of the thesis are as follows: to investigate the extent to which children's L2 acquisition is affected by the knowledge of L1, the extent to which frequency of words and word combinations in adult speech is reflected in the patterns of use in children, and the extent to which children

generalize knowledge between word forms or between constructions involving these forms in the process of developing productive constructions.

The linguistic phenomena chosen as the focus for the thesis are definite and indefinite articles and auxiliary verbs BE and DO. These aspects of the English grammar were chosen for several reasons. Articles are functional words with quite complex semantics that can influence the meaning of a phrase or sentence in subtle ways. Thus, articles are not straightforward to learn and previous research shows that English L1 and L2 learners have difficulty with articles and acquire them gradually. However, not enough is known about child L2 acquisition of articles. In the case of auxiliary verbs, the rules of form choice and positioning in a sentence are quite complex, and they play a role in the morphological form of the sentence (auxiliaries have to be correctly inflected) as well as in the syntactic form (auxiliaries have to be appropriately placed). By investigating the children's knowledge of the articles and the auxiliary verbs, we can get a more complete picture of acquisition of both the nominal and the verbal domains, thus filling the gap in the empirical knowledge of L2 development in children.

1.5 How this thesis is organized

This thesis is organized into 9 chapters. Chapter 2 discusses the theoretical issues raised in this chapter in more detail, evaluating each of them from the point of view of both the generative and the emergentist framework, and provides the context for the empirical chapters. Chapter 3 focuses on the linguistic structures under investigation, namely the system of articles of English and auxiliary verbs BE and DO in English. First, a linguistic description of each English structure is given. Then, previous acquisition studies of articles and auxiliaries are reviewed, followed in turn by predictions for children learning English as a L2, formulated in the generative and emergentist frameworks. Chapter 4 describes the method of obtaining and analyzing data used in the studies of articles and auxiliaries. It details the selection of participants and their background, and the types of speech elicitation tasks used in the studies. This chapter also explains the choice of adult

speech corpus that will be used as the model of input that the children receive. The two chapters that follow are results chapters focusing on the issue of L1 influence on the children's acquisition of English articles (Chapter 5) and auxiliaries (Chapter 6). Chapters 7 and 8 are results chapters dedicated to the analysis of these linguistic structures in the emergentist framework. While Chapters 5 and 6 will be concerned with group comparisons based on the children's L1 type, Chapters 7 and 8 will include analyses of longitudinal data from individual participants as well as overall generalizations about children's L2 development over time. The results in each of these four chapters will be tested against the predictions of each of the theoretical models formulated in Chapter 3. Finally, Chapter 9 will bring the findings together and discuss the implications of the empirical findings for the generative and emergentist models of acquisition.

CHAPTER 2. THEORETICAL APPROACHES TO LANGUAGE ACQUISITION

2.1 Introduction

This chapter reviews the models of language acquisition that will be tested using the data from L2 children. As previously mentioned in Chapter 1, the issues tackled in this thesis are L1 transfer in child L2 acquisition, the role of input frequency, and the development of productivity. In this chapter, two influential generativist models of L2 acquisition will be reviewed (Full Transfer/Full Access and feature assembly), as well as the usage-based model of acquisition. Each of the three issues pertaining to child L2 will be discussed from the perspective of these theoretical models. However, some of the models are better at interpreting certain aspects of L2 acquisition than others. Thus, mechanisms of L1 transfer have been better developed in the generativist models, so this issue will be discussed in more detail from a generativist perspective, with implications for the emergentist theory. On the other hand, constructivist models are centered on the role of input frequency and emergent productivity, so these issues will be discussed from the point of view of constructivism, with implications for the generativist theory.

2.2 Generative approaches

2.2.1 Full Transfer / Full Access

From a Universal Grammar (UG) perspective, L2 learners have either direct access to UG or indirect access via the working example of the grammar that their L1 provides. The approach that is based on the direct access option is one of the most influential views of L1 transfer in the generative framework: the Full Transfer/Full Access model (Schwartz & Sprouse, 1996). This model hypothesizes that the initial state of L2 acquisition is the entire grammar of the learner's L1 (full transfer) and that in the process of learning the grammar gets restructured, never violating the universal principles (full access to Universal Grammar). The restructuring of the initial L1 grammar happens when it fails to

account for the L2 input that the L2 learner receives. In the process of restructuring, learners develop new hypotheses that are constrained by UG. All of UG principles, whether or not they are present in L1, are available to L2 learners, in the same way as they are to L1 learners.

Some researchers argued in favor of a different model that assumed partial L1 transfer (e.g. Vainikka & Young-Scholten, 1996). They hypothesized that at the very early stages of L2 acquisition, learners are able to transfer the knowledge of lexical categories such as Noun and Verb into their L2, but do not transfer functional categories such as Tense and Agreement. Even though the authors found some empirical evidence for partial transfer, the most supported model to date is the full transfer model.

For example, learners whose L1s have functional projections of Tense and Agreement are expected to transfer these projections to L2. This will result in learners using agreement morphemes, auxiliary and modal verbs (i.e. morphemes that mark Tense and Agreement) in most of their sentences early on. Among the studies that argued for transfer of Tense and Agreement (together referred to as Inflectional projection) from L1 were Grondin and White (1996), Haznedar (2001, 2007) and Lakshmanan (1993/1994). Grondin and White found correct inflections, case marking, subject clitics and subject placement in the speech of English children learning French, interpreting it as evidence for the transfer of Inflection projection from L1 English into L2 French. Haznedar found evidence for transfer of Inflection from L1 Turkish into L2 English, indicated by very early use of copula BE. Lakshmanan analysed the speech of a Spanish-speaking girl learning L2 English and similarly argued that the presence of lexical material such as copula and auxiliary BE and infinitival to indicated the presence of Inflection in Marta's L2 English. The child used uncontracted forms of auxiliary and copula (e.g. ***Is*** *no going to rain there in Puerto Rico*) and the infinitival *to* (e.g. *This girl is* ***to*** *wash your hand*). According to Lakshmanan, it demonstrated that she transferred finiteness marking, which is located in the Inflection projection in sentence structure, from Spanish to English.

A specific version of the FT/FA model has not been developed for child L2. But since researchers argue that at the onset of L2 acquisition (5-6 years) children already have the knowledge of L1 in place (e.g. Schwartz, 2004; Unsworth, 2005), the predictions of the model has been applied to L2 children and adults equally. Just like adults, they are expected to transfer the entire clause structure for both declaratives and questions. Children whose L1s instantiate the abstract categories of Tense and Aspect will transfer it to their L2 English at the initial stage of L2 acquisition. Children whose L1s do not have these abstract categories initially will not be aware that it is obligatory and will omit auxiliaries. Gradually they will reconstruct their mental representation of the clause structure.

Generative theories of adult L2 acquisition disagree on the issue of the availability of 'new' abstract categories to a L2 learner. While some argue that any new category can be learned in a L2 due to full access to UG, others (e.g. Hawkins & Chan, 1997; Hawkins & Hattori, 2006) argue that the L2 categories that are not instantiated in L1 cannot be learned and that in such a case an L2 learner does not achieve a native-like competence of the abstract categories of L2. However, this is the case with adult L2 learners. In the FT/FA perspective, child L2 learners are expected to acquire new L2 categories successfully and achieve a native-like competence.

Empirical evidence for L1 transfer in L2 acquisition

The role of L1 in child L2 acquisition is an area of investigation that still has not generated a clear conclusion in previous research. While some researchers found no conclusive evidence for L1 influence on L2 children's accuracy with verb inflections (Dulay & Burt, 1974; Ionin & Wexler, 2002; Paradis, 2005), others reported L1 influence effects in several domains of L2 morphosyntax (Haznedar, 1997, 2001; Lakshmanan, 1995; Unsworth, 2005; Whong-Barr & Schwartz, 2002). For instance, in the area of child L2 acquisition of articles, Chondrogianni (2008) found evidence for the initial transfer of L1 properties in Turkish learners of Greek. However, Chondrogianni focused on the acquisition of only the definite article in comparison with pronominal clitics.

The studies that focused on the acquisition of auxiliary BE by L2 children are Gavruseva (2008) and Ionin (2008). Both authors suggested L1 transfer as a potential explanation for the patterns they found in their studies, but the type of data in these studies did not allow the authors to make definitive conclusions. Ionin's study had participants from only one background, namely Russian; Gavruseva's study had very few participants and potential L1 transfer was observed in only one child. In order to test for L1 transfer, it is necessary to compare two or more groups whose L1s contrast with English with respect to the organization of their auxiliary verb systems.

2.2.2 Feature assembly

The FT/FA model analyzes L1 transfer as transfer of abstract categories such as Determiner, Tense, or Aspect, rather than specific lexical items. This analysis can be too general for some cross-linguistic contrasts. For instance, one might ask whether differences are to be expected between groups of learners whose L1 article systems are parallel to English (i.e. they have a definite and an indefinite article) and those whose L1s are not completely similar (e.g. Arabic has a definite article but no indefinite article). Such differences cannot be captured by the FT/FA model since both English and Arabic presumably instantiate the grammatical category Determiner. Different modules of UG might play different roles in L2 acquisition. For instance, UG can be split into lexical and functional categories, which in turn contain functional features. A number of acquisition studies addressed the issue, examining whether functional features are available or not (Clahsen, 1996; Herschensohn, 2000; Hawkins 2001; Hawkins & Chan, 1997; White, 2003).

A feature-assembly approach (Lardiere, 2009) can better capture cross-linguistic differences such as the one between Arabic and English. The basic postulation of this approach is the same as that of the FT/FA model, namely that the fundamental functional categories are the same across languages. In addition, the feature-assembly approach provides a more detailed formal description of

differences among languages in terms of what features they select from the universal inventory and use in the assembly of functional categories.

This approach interprets the process of L2 acquisition as follows. Learners start the task of L2 acquisition equipped with the knowledge of the combinations of features and the mapping of these combinations onto morphophonological forms that are specific to their L1. According to Lardiere (2009), features that are not present in the L1 are in principle available to learners and acquirable, but morphological differences in how features are assembled in lexical items present a true learning problem, even in the case of L1-L2 pairs when both languages select the same subset of relevant features. The acquisition problem in this case involves the learners' figuring out how the relevant features are re-mapped onto new language-specific morphophonological forms.

The difference between L1 and L2 acquisition is that in L1 acquisition, all functional categories are assembled feature-by-feature, for which relevant features have to be extracted from the input mappings of phonological forms onto feature bundles. In L2, this process takes place for particular functional categories not available in the learner's L1. The feature-based approach can account for potential unequal difficulty in the acquisition of superficially similar morphemes that differ in the complexity of their feature composition. For instance, the indefinite article can only be inserted in [-definite, +singular] contexts, while the definite article can be inserted in any [+definite] context, which makes the indefinite article more featurally complex and thus potentially harder to acquire.

2.2.3 The issues of input frequency and early productivity in generativism

The FT/FA model and the feature assembly model are based on the generativist assumption that learning is constrained by UG principles. But even in this view, acquisition does involve learning language-specific information, i.e. the lexicon, and mapping specific words of a language onto categories known to the child from UG. Regardless of the theoretical approach, language acquisition is the result of exposure. However, according to the generativist view, external conditions such as input frequency are not viewed as the source of knowledge

since the child does not need them to capture abstract principles of grammar. External conditions relating to exposure to language are considered necessary only for triggering effect, i.e. for activating the knowledge of abstract categories that are instantiated in the given language. Thus, the frequency of particular words in the input is not considered an important factor in generativist approaches, in line with the Chomskyan tradition (Chomsky, 1965). However, not all generative linguists have rejected the importance of frequency in acquisition. For instance, Ligthfoot (1991) considers input frequency to be a major factor in diachronic language change. In recent studies, generative researchers have begun to give more consideration to the role of input frequency (e.g. Anderssen & Westergaard, 2010; Kupisch, 2007; Valian, 2009; Valian, Solt & Stweart, 2009; Anderssen & Westergaard, in press). These researchers suggested a learning mechanism that incorporates distributional regularities, especially frequent regularities, into the process of associating individual words with innate abstract categories. Importantly, they still deny that categories are *created* from exposure to individual specific words in the input and that the frequency of these words determines *how* categories are created. In contrast, constructivist models of language acquisition, which are reviewed in the following section, assign a critical role to the precise distribution and frequencies of forms in the input.

Regarding the issue of early productivity of children's utterances, maturational generativist models (e.g. Wexler, 1998) argue that the knowledge of some functional categories 'matures' when children are developmentally ready. In contrast, continuity models assume that children's competence is adult-like from the outset since they have access to all categories, lexical and functional, from the earliest stages of combinatorial speech. Regardless of the particular approach in generativism, since children have access to (at least some) abstract categories, their utterances are considered to be generated by productive morphosyntactic rules at the outset of combinatorial speech. In this perspective, children do not go through intermediate stages of local non-syntactic combinatorial rules that can be later abandoned or generalized. However, this is precisely the process that underlines acquisition from an emergentist perspective.

2.3 Emergentist approaches

The emergentist framework is a family of approaches that share several key insights. These approaches consider constructions – conventionalized pairings of form and meaning – to be the basis of the study of grammar. In addition, emergentist approaches emphasize that in the process of language acquisition, the grammar is constructed on the basis of the input, under the influence of general cognitive, pragmatic, and processing constraints.

In this view, constructions cut across all levels of grammatical analysis, including morphemes, words, word combinations, simple and complex sentences, and idioms of various kinds. Constructions are usually classified on the basis of the variability of each of their parts. A fully lexically specified construction can consist of one simple word or, in case of idioms, of several fixed words. Examples of such constructions would be words, such as *avocado* or *and*, and idioms with no variable parts, such as *thank you* or *you are welcome*. Fully lexically specified constructions are also called *fixed constructions* or *formulas*. Partially lexically specified constructions are combinations of morphemes or words that have variable parts, as well as parts that are filled. For instance, the progressive verb form *V-ing*, the plural form *N-s* are examples of partially specified constructions. Phrase- and sentence-level constructions such as *the X-er the Y-er* or *pull <someone's> leg* are also partially specified. Such constructions are also called schemas or frames with a slot, with the slot being the variable, or lexically unspecified, part of the construction. The more variable slots a construction has, the more productive it is. Thus, fully lexically specified constructions are unproductive. The most productive constructions with all slots lexically unspecified are, for instance, the ditransitive construction *N V N1 N2*, or the passive construction *N aux VParticiple (by N)* (Goldberg, 2006: 5).

As an illustration of construction types, consider the "incredulity construction" example from Tomasello (2000: 236):

(1) My mother ride a motorcycle!
Them come to the party!

This construction can be considered non-canonical in comparison with the most typical form of an English sentence. However, it is highly abstract in the sense that it is not dependent on any particular word or phrase, and it is highly productive in the sense that all the slots in this construction can be substituted with a variety of words, producing a number of novel utterances. An example of a partially lexically specified construction would be the *let alone* construction (Tomasello, 2000: 237)

(2) I wouldn't live in Boston, let alone in New York.

This is a "mixed"' construction, that is, a construction that is in some parts abstract but that revolves around particular lexical items. In other words, constructions can be set phrases or abstractions that range over many specific phrases. Even the most productive constructions with variable slots can be productive to different degrees. For instance, only a limited number of verbs can fill the slot in the ditransitive construction, and the slot in the plural N-s construction can only be filled by countable nouns.

Since constructions cut across different levels of grammatical analysis, advocates of this approach argue that there is no discrete cutoff point between lexicon and grammar (e.g. Bybee, 2007; Croft, 2001; Langacker, 1987). Adults are assumed to store individual words, idioms, and partially specified constructions for use on a specific occasion, as well as more productive constructions. Emergentist approaches do acknowledge the existence of abstract grammatical categories such as noun, verb, or auxiliary, but they are thought to exist only to the extent that they can be derived from the distributional regularities of the input, which is in contrast with the generative approach that postulates that abstract grammatical categories are innate. Another contrast between the two approaches is that the emergentist approach, rather than assuming that speakers generate utterances by applying rules to appropriate grammatical categories, argues that speakers retrieve partially or fully lexically specified constructions

from memory, filling the slots in constructions and combining them. The generative approach does acknowledge that that some multiword utterances are stored in the lexicon, but limits it to idiomatic expressions. In contrast, the emergentist approach postulates a continuum of structure types stored in speakers' memory that ranges from individual words to multiword constructions, which can be fully or partially abstract. Finally, emergentist theories do not postulate any empty elements in language structures, as they do not derive one construction from another. An actual utterance or phrase is seen as a combination of several different constructions. For instance, the structure of a NP can be analyzed on different levels of abstraction (Bybee, 2007: 325):

(3) a. Very specific: *my mother*, *my computer*, *the car*, *a problem*, *an idea*
b. Partially general: [my + NOUN], [POSS PRO + mother]
c. More general: [POSSESSIVE + NOUN]
d. Fully general: [DETERMINER + NOUN]

In order to clarify what is meant by a construction on sentence level, consider the example of the ditransitive construction (Goldberg, 2006: 20). The meaning-surface form paring for this construction is represented as follows:

(4)

Semantics:	Intend-CAUSE-RECEIVE	(agent –	recipient –	theme)
Syntax:	verb	Subj	Obj1	Obj2

The first line represents semantic roles of the elements of the construction. Importantly, terms such as agent, recipient, etc., are used by Goldberg simply as labels for convenience, without any assumption about universality of these roles. The surface form of some constructions does not need to specify word order or even particular grammatical categories. For instance, some constructions will not specify an object of a verb, but the ditransitive construction does. It is important

to realize that the syntactic part of the construction is an abstraction from specifics of surface form, which can consist of a combination of several constructions. Thus, each element labeled as *subject*, *object*, and *noun* in (4) can be in turn analyzed as a construction its own right. But since there are certain generalizations that can be made about these slots in the construction (e.g. their typical meaning and position in the sentence), it is convenient for researchers to refer to them with these labels.

2.3.1 The usage-based model

The emergentist view of language and acquisition outlined above represents a considerably diverse set of approaches united by the support of the idea of emergent grammar and opposition to the UG-based theory. This thesis will mainly focus on one member in the family of emergentist approaches, namely the usage-based model (Bybee, 1995; Goldberg, 1995; Langacker, 1988). This model emphasizes the role of experience in the formation of linguistic categories and structures. It is a model of language in which grammar is seen as emerging from the patterns of language use in discourse. In this view, grammars record facts about the actual use of specific instances of linguistic expressions, such as their frequency and function, together with generalizations that arise from the repetition of linguistic expressions. This system is known as an emergent or usage-based system. In such a view, repetition is a factor in the formation of constructions because the human mind is sensitive to repetition in general, not only in the case of language, but also in the case of other cognitive domains (Bybee, 2007: 8). Thus, frequency of language expressions plays a very important role in the usage-based approach to grammar.

2.3.2 Input frequency in the usage-based model

There are various ways of counting frequency and various types of units that can be counted. Bybee (2007) distinguishes token and type frequency in her investigation of frequency effects. Token frequency is the number of times a unit appears in discourse. This unit can be a word, such as *are* or *the*, a word combination, such as *I am*, or a sentence, such as *I dunno*. Type frequency refers

to the number of distinct items that can occur in the open slot of a construction or the number of items that exemplify a pattern. Type frequency can apply to morphological patterns, such as the English regular past tense costruction V-*ed*, or to syntactic patterns, such as the ditransitive construction. Type frequency is one of the factors that determines the productivity of a construction. A highly productive construction is a construction that applies to a high number of distinct items and can also be easily applied to new items with a similar meaning or function. For example, the English ditransitive construction N V N1 N2: *He gave her the ball* is less productive than the N V N2 to N1: *He gave the ball to her* because the latter pattern is possible with a larger set of verbs (Goldberg, 1995).

Evidence for the role of input frequency

The usage-based view is supported by recent research that suggests that children's early grammatical knowledge may initially be tied to particular lexical items and only gradually becomes more complex and adultlike (Tomasello, 1992; Pine, Lieven & Rowland, 1998). Furthermore, children's knowledge of individual lexically based constructions and the development of more abstract constructions is thought to reflect the specific properties of the language they hear. Recent research suggests that relationships exist between children's early language use and the language they hear when examined at the lexical level, providing support for this approach.

Psycholinguistic research demonstrates that language processing remains sensitive to frequency effects even in adults (Bybee & Hopper, 2001; Ellis, 2002), but the majority of frequency-based studies have been dedicated to children acquiring L1. A growing number of studies provide evidence for the role that input frequency plays in children's language development (Cameron-Faulkner, Lieven & Tomasello, 2003; Matthews, Lieven, Theakston & Tomasello, 2005; Rowland & Pine 2000; Rowland, Pine, Lieven & Theakston, 2005; Theakston, Lieven & Tomasello, 2003, among others). For auxiliary verbs in particular, frequency effects have been found in studies of children's production and comprehension (Theakston, Lieven, Pine & Rowland, 2005; Theakston &

Rowland, 2009a,b; Wilson, 2003). It has been found that forms (both simple words and multiword constructions) that are highly frequent in the input are among the earliest to appear in child speech. However, as mentioned earlier in this section, both type and token frequency have to be taken into account. For instance, Bybee & Scheibman (1999) suggested that high token frequency and low type frequency of phrases results in entrenchment or chunking. For example, frequent occurrence of *Is it…?* as opposed to other questions with BE such as *Are they…?* in the input can result in the child using *Is it…?* as a universal interrogative marker in all contexts. By contrast, high type frequency and low token frequency result in productivity. In child speech, this can be the case with the transitive construction N1 V N2.

It has been demonstrated that L1 learners produce grammatical morphemes more accurately in constructions they have heard them in more often (Gathercole, Sebastián, & Soto, 1999; Pizzuto & Caselli, 1992; Tomasello, 2003). Opponents of the frequency-based account pointed out the fact that L1 and L2 learners have difficulty with many frequent morphemes, e.g. third person singular *–s* or articles (Ellis, 2002; Gass & Mackey, 2002) and with frequent sentence types, e.g. questions (Mackey, 1999). In defense of the frequency-based account for L1 (Lieven & Tomasello, 2008) and L2 acquisition (Ellis & Larsen-Freeman, 2006; O'Grady, 2008; O'Grady et al., 2009) researchers argued that in addition to frequency, other factors need to be taken into account, e.g. how input frequency interacts with children's current knowledge of lexically based or more abstract constructions, their sociocognitive interests and the semantic complexity, functional transparency, and phonological salience of the forms. Although type and token frequencies in the input are seen as central in the acquisition process, exactly how children's constructions develop into more abstract schemas is not fully understood. development of more abstract constructions from simpler ones may be at least partially driven by children gaining a fuller understanding of the semantics of more lexically specific constructions.

2.3.3 Early productivity in the usage-based model

In the usage-based approach to L1 acquisition, constructions are understood to be learned on the basis of the input and general cognitive mechanisms, such as intention-reading, analogy, entrenchment, and distributional analysis (Tomasello, 2003). This view of language acquisition resulted in research that focuses on specific words and word combinations that are produced early by children and how these early combinations begin to pattern into constructions, and how constructions are organized in a structured inventory (Dabrowska, 2000; Lieven, 2009; Dabrowska & Lieven, 2005). In the adult inventory, constructions exist at all levels of abstractness and complexity, and they are inter-related to form a dense network, which is not the case with children's inventory. Language acquisition is understood as the process of gradual building of the network of constructions, and it includes the following phases of development. The child starts off with individual words and fully lexically specific combinations, goes on to build partially lexically specified constructions that are more productive than fully fixed constructions, and finally generalizes across specific constructions to create more complex constructions (with more parts) and more abstract constructions (with more variable slots). It has long been recognized in L1 acquisition research that at the earliest phase of language development, long strings composed of several words can be treated as a single unit by the child. For instance, Brown (1973) noted that some of his subjects' sentences were memorized wholes. He hypothesized that prefabricated routines such as *what's that* were the result of very high input frequency of a structure that was perceptually salient and that was, at that time, beyond the child's linguistic maturational level. A large body of recent work has confirmed that the child's early knowledge of grammar is tied to individual lexical items or lexically-specific constructions (e.g. Dabrowska & Lieven, 2005; Rowland & Pine, 2000; Theakston et al., 2005, among others). While the mainstream generative approaches consider memorized word combinations separate from children's knowledge of a rule-governed system that is present from birth, the usage-based approaches argue that language acquisition is piecemeal: learners do not

instantaneously acquire an entire abstract category or rule that applies across the board to all members of that category or to all structures defined by that rule. Instead, learners can gradually become selectively productive with some members of categories and some structures, while other members of the category or structures remain unproductive. In other words, "rules" are emergent general or abstract constructions and an abstract construction is the end point of learning, not the beginning, in contrast to generative theories. Language learning progresses from memorized chunks of one or two words to more complex phrases that have slots into which the child can place a variety of words (e.g., *I don't* + Verb; *where's* + Noun?). As children recognize the substitutability of various elements in constructions, constructions become more and more productive, i.e. applied to more and more new items. Finally, children generalize across constructions, creating more abstract ones.

With respect to morphology, usage-based accounts propose that learners acquire and store morphemes as part of their initially unproductive constructions. Again, as the acquisition proceeds in a piecemeal fashion, differentiation of individual morphemes with their interconnected semantic and phonological features happens gradually over time in children's lexicons (e.g., Bybee & Hopper, 2001; Ellis, 2002; Tomasello, 2003). Thus, under a usage-based approach to language, there is a continuum of construction types that speakers operate with. It includes two clearly opposite kinds of structures, fully lexically specified constructions and fully abstract constructions, as well as constructions with specified and open slots that can be seen as occupying the space between the two extremes of the continuum. Usage-based studies of language acquisition are aimed at tracing how these early constructions become more abstract and part of a wider group of interconnecting constructions.

Routines and productivity in child L2 learners

All the existing studies of learners' productivity with auxiliaries have been concerned with L1 acquisition. The usage-based framework has not been applied to L2 children in the way it has been used in L1 studies of Tomasello, Lieven and

colleagues cited above, but some less theory-driven work has been done on formulaic language in this learner population, reviewed in detail in Wray (2002). Comparing L2 children to L1 children, we need to take into account the fact that L2 children are more cognitively and linguistically mature at the onset of acquisition. Their memory capacity is greater, and they have established a substantial knowledge of L1 grammar when they start to learn a L2.

Based on several studies of L2 children from different backgrounds, Wray (2002: 166) described the process of the development of productive language in L2 children as follows. The acquisition pattern starts with some sort of silent period that is longer for some learners than others. It is followed by a period of imitation without full comprehension. Next is a period of apparent competence, characterized by fluent and accurate output made up of formulaic sequences used in pragmatically appropriate ways. The following are some of the typical memorized expressions found in the spontaneous speech of Spanish-speaking children learning L2 English (Wong Fillmore, 1979: 211):

(5) *I don't care.* *Do you wanna play?*
I dunno. *Waddya wanna do?*
You know what? *I'm gonna tell on you.*

Several studies have suggested that L2 children rely more on routines and patterns than do L1 children for various reasons (Hakuta, 1974; Hatch, 1972; Krashen & Scarcella, 1978). Hatch (1972) suggested that, in comparison with L1 children, L2 children use more memorized routines in their speech due to their ability to store and repeat longer imitated sentences. Hakuta (1974) also emphasized that L2 children may rely on memorized segments more than L1 children because of their greater need to communicate and because of the higher semantic complexity of the information they need to convey. In a usage-based view of L1 acquisition, the need to communicate is what drives L1 acquisition in general and the acquisition of conventional units such as those given in (2) in particular (e.g. Tomasello 2000). This need to communicate is even greater for L2

children (due to the pressure to interact in the classroom, for example), and the absence of sufficient L2 knowledge may result in an increased need to use conventional units, supported by an advanced ability to pick up and memorize the necessary formulas.

There is individual variation reported in children's learning styles, but similar formulaic constructions were found across different children (e.g. Hakuta, 1974; Tabors, 2007; Wong Fillmore, 1979). At the time when children heavily rely on such utterances they may sound fluent but they make little attempt to dissect these utterances. Subsequently, when the children attempt to express novel ideas, their output becomes less fluent and native-like. For instance, the Japanese-speaking child studied in Hakuta (1974) used what appeared to be a grammatical *how-to* construction in such utterances as *I know how to do it* and *I know how to make*. However, the construction disintegrated over time, resulting in errors, for instance: *We only know how do you make it like that* and *I know how do you make it*, which possibly resulted from the child misanalyzing the structure as *I know* + *wh*-question. In other words, as control of the language increases, the formulaic sequences are used more creatively. They are added together or embedded into novel structures to make new utterances. Finally, formulaic sequences can be segmented into smaller strings. According to the usage-based view, this process of increasing creativity is at the heart of language acquisition process.

The development of fully abstract constructions is thought to depend on input frequency, but children's growing awareness of both the structural and semantic similarities between constructions also needs to be taken into account. Abstraction may be at least partially driven by children gaining a fuller understanding of the semantics of lexically specific constructions. Older children learning an L2 are expected to be faster at gaining such an understanding. Children are expected to have a greater memory capacity to memorize longer utterances by the time they start acquiring L2, thus it is not surprising that the studies mentioned above found formulaic usage of sentence-level constructions. So it could be the case that L2 children start off with fully lexically-specific chunks that are longer than those found in L1 acquisition, but due to more

developed cognitive capacities and experience with learning their L1, they might arrive at abstract constructions faster.

2.3.4 L1 transfer in the usage-based model

All theoretical approaches to L2 acquisition agree that L1 transfer impacts acquisition in both positive and negative ways. As mentioned in Chapter 1 (Section 1.3), proponents of the usage-based account acknowledge that frequency effects cannot be considered in isolation and that the effects of L1 transfer need to be taken into account, in addition to such factors as perceptual salience and semantic and morphological complexity (Ellis, 2002; Gass & Mackey, 2002). The mechanics of L1 transfer have not been formalized in the usage-based model. For instance, the discussion of the role of L1 in L2 acquisition in Bybee (2008) is limited to the following statement:

> To the extent that the constructions in the second language are similar to those of the first language, the L1 constructions can serve as the basis for the L2 constructions, with only the particular lexical or morphological material changed. However, since even similar constructions across languages are likely to differ in detail, the acquisition of the L2 pattern in all its detail is hindered by the L1 pattern. (p. 232)

This line of thought is similar to the idea that children's growing awareness of syntactic similarity between constructions needs to be taken into account in the analysis of both L1 and L2 development. It was discussed in Gathercole (2007) for bilingual acquisition, and in Dabrowska (2000) and Siebenborn (2010) for monolingual acquisition. In Gathercole (2007), a step forward was made in developing a constructivist view of cross-linguistic influence in bilingualism. She suggested that, in order for language transfer to occur, "the child has to have discovered some general pattern across a range of linguistic items" and "the patterns in question will have to be abstract enough that they do not rely on lexically specific information" (Gathercole 2007: 240). In other words, in order to explain L1 transfer in terms of a constructivist approach, we need to suppose that

some generalized schemas, i.e. the most prototypical constructions, are abstract enough so that they can be stored separately from their specific instantiations. This way of looking at cross-linguistic influence was suggested by Gathercole for bilingual acquisition. However, it could be even more applicable to child L2 learners than to children learning two L1s simultaneously. At the onset of L2 acquisition, L2 learners already have a substantial knowledge of another language, and so they may be more likely to have established some constructions in their L1 that are abstract enough to be transferred to L2.

An important question for L2 applications of the usage-based model is how L2 learners become aware of similarities between constructions and if so, at what level of abstractness the similarities can be observed. Given the broad range of phenomena encompassed by the notion 'construction' (recall example (1) in Section 2.3 that listed the constructions involved in a noun phrase), the usage-based model needs to be more specific about defining construction similarity. The concept of construction similarity is related to another important aspect of construction learning, namely syntactic complexity, i.e. the number of slots a construction has. While it is uncontroversial that learners need to form analogies between similar instances of constructions in the input in order to make a low-level generalization, the analogy process has been underspecified. One exception is a study investigating the relation of known constructions to the learning of new ones by Abbot-Smith & Behrens (2006), who showed how the acquisition of the German passive could be supported by the knowledge of related constructions involving the verb *sein* 'be' or past participle. The idea that children use already known constructions to learn the form-function mappings of new constructions within their L1 will be also relevant to the question of whether L2 learners are able to apply known constructions from their L1 to deal with construction-building in a L2. A usage-based interpretation of L1 transfer in terms of syntactic constructions will be discussed in the last chapter of the thesis (Section 9.6).

2.4 Summary

The main feature of the FT/FA and feature assembly model discussed in this chapter is the abstract knowledge of functional categories that is available to L2 learners. Since the two generative acquisition models are well suited for making explicit predictions regarding the transfer of L1 categories or features into L2, the analyses of L1 effects in child L2 in this thesis will be based on the predictions of these two approaches. As mentioned in Chapter 1 and in the previous section, L2 acquisition in general and the issue of L1 transfer in particular have not been systematically addressed in usage-based studies. While researchers working in the usage-based framework do not deny the possibility of language transfer in bilingual L1 acquisition (Gathercole, 2007) as well as in L2 acquisition (Bybee, 2008), the mechanics of the transfer process and the potential extent of L1 influence have not been formulated in this framework, possibly because this theory has been developed most intensively for monolingual L1 acquisition. Chapters 5 and 6, dedicated to the effects of L1, will focus on testing the predictions of the UG-based accounts. However, after considering child L2 data from a usage-based perspective in Chapters 7 and 8, an alternative to the generativist account will be put forward, in an attempt to work out the possible mechanics of L1 transfer from the usage-based perspective, motivated by observed effects of input frequency and emergent productivity in child L2 learners.

CHAPTER 3. ASPECTS OF GRAMMAR IN QUESTION

3.1 Introduction

The auxiliary verb system and the article system in English are complex and were shown to pose problems for L2 learners of English. These two domains can be a potential source of difficulty in child L2 learners, and they are also expected to be influenced by L1 transfer in certain groups of L2 children, depending on the typology of their L1. Exactly these effects will be investigated in terms of the three approaches discussed in Chapter 2. First, the system of articles in English are briefly described (Sections 3.2.1-3.2.2), followed by a literature review of previous acquisition studies of articles in Section 3.2.3. In Section 3.3, all L1s of the children in the study are contrasted with English in terms of presence or absence of articles or morphemes with similar grammatical functions.

The second half of the chapter has a similar structure and discusses the auxiliary system of English. The auxiliary verb system of English is described in terms of the generativist and the constructionist approaches (Sections 3.4.1-3.4.2), followed by a discussion of previous studies of the acquisition of auxiliaries in English (3.4.3). Section 3.5 describes the counterparts of the English auxiliary system in the four languages represented in the backgrounds of the participants and discusses the similarities and differences of these languages and English. Finally, in Section 3.6 expectations are formulated for (1) the cross-group comparisons of L1 effects in the acquisition of articles and auxiliaries by L2 children, and (2) for the patterns of development of learners' productivity in the two domains, in terms of the generativist and constructionist approaches to acquisition discussed in Chapter 2.

3.2 The article system of English

Frequency counts of the Corpus of Contemporary American English (COCA; Davies, 2008) reveal that articles are one of the most frequent words in English. In fact, *the* is the most frequent word (over 22 million tokens in the 400-

million word corpus), and *a* occupies the 5th place (over 10 million tokens). Despite their frequency in speech, articles remain to be one of the most problematic areas for L2 learners, perhaps because of the variety of contexts each article can be used in and because of the absence of clear-cut distinctions among the conditions that determine article choice. In addition, articles do not have clear referents and might be hard to perceive because they are less acoustically salient.

3.2.1 Articles in the generativist perspective

In the generativist view, articles are exponents (phonological representations) of the functional category Determiner. Determiners other than articles can occupy the head position in a DP, such as for instance possessive pronouns or demonstrative pronouns. Articles are considered a special kind of determiner because it is a specialized grammaticalized means of encoding the semantic contrast of definiteness (Lyons, 1999).

There are many ways to classify the English article system. One classification system used in several studies of child and adult acquisition of English, such as Lu (2001), Robertson (2000), and Thomas (1989) comes from Bickerton (1981). Under Bickerton's analysis, determiners are classified as combinations of semantic features [±specific reference] (SR) and [±hearer knowledge] (HK). An NP has the feature [+specific reference] when it refers to an entity that exists but cannot be identified by the hearer from the previous conversation or from the situation. An NP has the feature [+hearer knowledge] when the hearer knows which entity the NP refers to from what has been said before or from the context. The interpretation of English articles in relation to [SR] and [HK] features was used in earlier studies (e.g. Huebner, 1985; Parrish, 1987; and Thomas, 1989), as well as in recent studies (e.g. Lu, 2001; Robertson, 2000). Not all researchers used the [HK] [SR] features as the basis for classification of articles. Ionin, Ko & Wexler (2004) used a slightly different classification or articles using [±definite] and [±specific] features, which is essentially the same as Bickerton's, because [±definite] feature can be interpreted as 'mentioned' or 'not mentioned', or [HK]. Definitions of what it means for a

noun to be definite or specific differ across studies, but the crucial semantic contrast grammaticalized in English is whether the referent is a part of the hearer's and/or the speaker's beliefs.

The feature assembly approach has often been used in recent studies on the acquisition of English articles (Hawkins et al., 2006; Lardiere, 2009; Snape, Leung & Ting, 2006; Wakabayashi, 2009). Recall that the essence of the feature assembly approach (Lardiere, 2000, 2004, 2009), which is a type of generative approach, is the assumption that the output of syntactic operations produces strings of terminal nodes that host clusters of morphosyntactic features. Lexical items (i.e. actual sound-meaning pairings) are inserted into terminal nodes after all syntactic operations have applied. According to Lardiere (2009), formal linguistic features reflect fundamental cognitive categories. While the fundamental categories are the same across languages, languages differ in terms of what features they select from the universal inventory and use in the assembly of functional categories. Applying a feature-assembly approach to article systems, it was argued that articles are exponents (phonological representations) of the functional category D, which can host clusters of formal features such as [definite], [specific], or [singular] (Lyons, 1999; Ionin, 2003; Ionin et al., 2004). On a feature-based approach, the definite and indefinite articles in English are inserted in the terminal nodes with the following bundles of features:

(1) [D, +definite, +singular] = 'the'
[D, +definite, -singular] = 'the'
[D, -definite, +singular] = 'a'

Combinations of these features and the projection of the functional category D describe the syntactic aspect of the article system.

An analysis of articles in terms of bundles of features can be applied both in L1 and L2 acquisition. According to the generativist approach, syntactic categories are innate and the general form of the grammar is innate. This claim implies that the child innately knows that a Determiner Phrase has a determiner as

its head and takes a noun as its complement. Under a feature-based approach to acquisition (Hegarty, 2005; Lardiere, 2009), there is a universal set of features available to the child, and the child's acquisition task is to select only those features that are deployed in L1, while disregarding other features. Selected features are assembled into language-specific lexical items that enter into the computation. Importantly, children acquiring their L1 compose lexical items feature by feature, since the universal set of features is freely available to them, but specific feature bundles are not. The situation is expected to be different for L2 children, who presumably have access to the set of features and who also might have experience assembling features into lexical items similar to English articles in their L1. Some of the children in the study come from L1s that do have article systems. For such learners, we can expect L1 transfer effects in the form of re-mapping known feature bundles onto new L2 forms, which is the case of L2 acquisition. This analysis is promising for the comparison of learners with L1 backgrounds that differ in the manifestation of the article system, for instance languages with no articles (Mandarin and Cantonese) compared with languages that have only one article (Arabic), in turn compared to languages with a two-article contrast (Spanish).

3.2.2 Articles in the constructionist perspective

The constructionist view of grammar (Croft, 2001; Goldberg, 1995) belongs to the set of emergentist approaches (Chapter 1, Section 1.2; Chapter 2, Section 2.3), which are juxtaposed with the generativist program in the present study. While the constructionist approach is a type of emergentism that focuses on describing language structures, the usage-based model (reviewed in Chapter 2.3) focuses on describing the process of acquisition of these structures. In the constructionist view, the structure of a phrase can be analyzed on different levels of abstraction. Consider the example of the noun phrase (Bybee, 2007: 325):

(2) a. Very specific: *my mother*, *my computer*, *the car*, *a problem*, *an idea*
b. Partially general: [my + NOUN], [POSSESSIVE PRO + mother]

c. More general: [POSSESSIVE + NOUN]

d. Fully general: [DETERMINER + NOUN]

In the process of acquisition of articles, the knowledge underlying young children's very early speech is not syntactic categories of the type in (2d) but rather specific or partially general expressions such as the ones in (2a) and (2b).

3.2.3 Previous acquisition studies

Numerous studies have been dedicated to the acquisition of article systems in L2 in general and in L2 English in particular. In L2 acquisition research there is agreement that all English language learners have difficulty in using articles (at least initially), but there is no consensus as to what the reasons for this difficulty are, and whether these reasons could be the same for all L2 learners. Most of the work on the acquisition of L2 English has focused on adult learners of English with varying degrees of proficiency. In addition, prior research has almost exclusively focused on L2 learners with L1s lacking articles, such as Korean, Russian, Japanese, or Mandarin Chinese (e.g., Ionin & Wexler, 2003; Ionin et al., 2004; Robertson, 2000; White, 2003). However, the only way of confirming L1 transfer effects is to compare two or more groups of learners from typologically different language backgrounds. A few recent studies (Hawkins et al., 2006; Snape et al., 2006; Ionin, Zubizarreta & Maldonado, 2008) contributed more to the discussion of the role of L1 in the acquisition of articles by comparing L2 learners from L1 backgrounds that contrast with respect to the presence of article systems similar to English. These studies found that L1 transfer facilitates L2 acquisition in learners from such language backgrounds as Greek or Spanish, i.e. languages that have the definite/indefinite article distinction, indicated by native-like performance of learners from such language backgrounds.

Research on adult L2 acquisition of articles has reported errors of omission and substitution. Article omissions were reported in learners of English whose L1s did not have articles (Huebner, 1985; Lardiere, 2004; Parrish, 1987; Robertson, 2000; White, 2003), and such learners' difficulties with articles were

naturally attributed to the influence of their L1s. In addition to omission, difficulty in choosing appropriate articles was also documented in these learners, who produced more errors of substitution than speakers of languages with article systems similar to English, such as French (Sarko, 2008), Greek (Hawkins et al., 2006), or Spanish (Ionin et al., 2008; Snape et al., 2006). Omission and substitution errors were documented in learners from no-article L1 backgrounds of all proficiency levels, even in learners who reached a stable state of L2 development (Lardiere, 2004; White, 2003). Various groups of L2 learners were consistently reported to be more accurate with the definite article than with the indefinite article (e.g., Lardiere, 2004; Robertson, 2000; White, 2003). In other words, learners seem to acquire *the* earlier than *a*. Lardiere (2004) suggested that L2 learners find it easier to acquire *the* because 'definite articles in English need not take number and the count/mass distinction into account, which makes them less featurally complex than indefinites in at least one respect' (p. 335). If featural complexity is the underlying reason for the acquisition order of *the* and *a*, then L2 learners are expected to display this pattern regardless of L1 background.

Support for the similarity of child L2 and adult L2 acquisition comes from the studies that found L1 transfer effects at the early stages of child L2 acquisition of syntax (Unsworth, 2005; Haznedar, 2001). In the area of child L2 acquisition of articles, Chondrogianni (2008) found evidence for the initial transfer of L1 properties in Turkish learners of Greek. However, Chondrogianni focused on the acquisition of only the definite article in comparison with pronominal clitics. Studies of L1 acquisition and L2 acquisition of English found similar types of errors in learners, namely substitution of the definite article in indefinite contexts and omission of articles. Previous studies seem to converge on the finding that L2 learners of English reach higher accuracy with the definite article than with the indefinite article at the early stages of acquisition. In a recent study of 10-12 year-old children acquiring L2 English, Ionin, Zubizarreta & Philippov (2009) found similar errors of article substitution in the 10-12 year-olds and adults learning L2 English. Even though the underlying reasons for the similarity of errors in learner populations could be somewhat different due to age of onset, there appears to be

continuity in the development of the article system in L2 acquisition in learners of all age groups.

3.3 Article systems in languages represented in the study

On the basis of the children's L1 characteristics in this study, in the first part of the study they will be split into two subgroups, [–article] and [+article] L1s simply based on whether or not the language has an article morpheme of any kind. The [–article] L1s will include Mandarin and Cantonese Chinese, Korean and Japanese. The [+article] group will include Arabic, Spanish, and Romanian. In the second part of the study, L1 transfer effects will be investigated further by comparing four separate L1 groups (Arabic, Chinese, Hindi/Urdu/Punjabi and Spanish) in order to find out whether specific similarities or dissimilarities between L1 and L2 affect learners' performance. This section provides details justifying this classification.

3.3.1 Arabic

There is a definite article in Arabic realized as the prefix *l-* or *al-* (Bateson, 1967). The definite article is prefixed to the noun and assimilated to certain initial consonants (e.g. *l+šams – ššams* 'sun' – 'the sun'). The word-final vowel marks case.

(3) **'al**-kitāb-u
the-book-nominative

In spoken Arabic, there is no marker of indefiniteness. Without the definite article, the noun is unambiguously indefinite, unless it is followed by possessive or pronominal determiners.

Since Arabic has a definite article, under a feature-based approach we can assume that it grammaticalizes the feature [definite]. Unlike in Spanish or English, the phonological exponent of the feature [-definite] is null in Arabic, which is similar to plural indefinite nouns in English, except in Arabic both singular and plural nouns are bare. In generative grammar terms, Arabic projects a functional category D.

In the first half of the study, Arabic speakers will be grouped together with Spanish speakers in the [+article] L1 group and will not compare children within the group. In the second half of the study, larger numbers of participants will allow me to separate the Arabic and the Spanish speakers into two groups in order to see whether there are more subtle L1 effects due to the differences in the article systems of these languages.

3.3.2 Spanish

The Spanish system of articles is similar to the English system (Butt & Benjamin, 1988): nouns can have a definite article, an indefinite article, or no article. In addition to definiteness, articles in Spanish also agree with the noun in gender and number. There are indefinite masculine and feminine articles *un/una*, and singular and plural definite articles *el/los* (masculine) and *la/las* (feminine).

(4)	a.	*un(o)* m.sg.	*unos* pl.sg.	indefinite articles
		una f.sg.	*unas* f.pl. 'some'	
	b.	*el* m.sg.	*los* m.pl.	definite articles
		la f.sg.	*las* f.pl.	

(5)	**Un**	libro	está	en	**la**	mesa
	a-masc.sg.	book	be-3sg.pres	on	the-sg.fem	table
	'A book is on the table'					

While the article systems in Spanish and English are very similar in terms of the definiteness/indefiniteness encoding, there are particular uses of articles that differ in the two languages. For instance, unlike in English, bare nouns in Spanish cannot occur before the verb, but they can occur in post-verbal positions.

(6)	a.	Llegaron	estudiantes
		arrived	students
		'(Some) students arrived'	

b. *Estudiantes llegaron
students arrived

Another difference between Spanish and English is the use of the definite article with generic nouns: *la naturaleza* 'nature', *la pronunciación española* 'Spanish pronunciation' Butt & Benjamin (1994: 22). Despite several idiosyncratic uses of articles, by and large it can be argued that the feature [definiteness] is grammaticalized in Spanish and that the article system of Spanish is parallel to English in terms of the indefinite/definite article distinction (see e.g. Zagona, 2002).

Rumanian was the L1 of one of the children in the longitudinal study. This language will be included in the [+article] group together with Arabic and Spanish, as its article system is very similar to Spanish with definite and indefinite articles that agree with the noun in gender (Mallinson, 1986).

3.3.3 Mandarin and Cantonese Chinese

These languages do not have counterparts of English articles (Li & Thompson, 1997; Matthews & Yip, 1994). Mandarin and Cantonese have a quantifier 'one' that can be used to mark indefiniteness, and demonstrative pronouns 'this' and 'that', which can express definiteness, similarly to English. According to Chen (2004), Mandarin Chinese does have various devices to indicate the universal pragmatic notion of identifiability but the grammatical feature [definiteness] does not yet exist in the language. As for Cantonese, Matthews & Yip (1994: 89) argue that the language has no articles equivalent to *a* or *the*. The word *yat* 'one' may be used like an indefinite article, referring to an indefinite object or person, but it is optional.

(7) (yat) **ga** che
one CLS[2] car

[2] CLS – classifier

Chen argues that demonstratives and the numeral 'one' are developing uses of definite and indefinite articles respectively, but morphologically they have not been fully grammaticalized, which appears to be the consensus in the theoretical studies of Mandarin and Cantonese (e.g., Chen, 2004; Cheng & Sybesma, 1999; Leung, 2005). The system of classifiers in Mandarin and Cantonese cannot be likened to the English article system, even though classifiers perform some of the functions of the English articles in individuating entities. For instance, in Cantonese, when the noun phrase is a subject or a topicalized object, the presence of a classifier denotes a definite person or object (Matthews & Yip, 1994: 89).

(8)	ga	che	jo-joyuh	**go**	cheut-hau
	CLS	car	block-PROG	CLS	exit-mouth

'The car is blocking the exit' Not 'A car is blocking the exit'

By contrast, a noun with classifier following the verb may be definite or indefinite.

(9)	ngoh	tingyaht	wuih	wan	**go**	leuhtsi
	I	tomorrow	will	contact	CLS	lawer

'I'll contact a/the lawyer tomorrow'

For purposes of the present study, it suffices to say that in Cantonese and Mandarin, the functions similar to those of the English articles are performed by a combination of devices, such as classifiers, numerals, and word order. Thus, unlike English, these languages do not have an article system, i.e. a set of morphemes that developed highly specialized uses to indicate (in)definiteness. In line with the conclusions of the theoretical studies mentioned in this section, the present study will assume that both Mandarin and Cantonese do not grammaticalize the feature [definiteness]. The same argumentation applies to

Korean and Japanese (Ionin, 2003; Kuribara, 1999), which will be classified together with Mandarin and Cantonese as the [–article] group of languages.

3.3.4 Hindi, Urdu and Punjabi

These three languages were grouped together for the purposes of this study because they are very closely related. In fact, differences between Hindi and Urdu are mostly sociolinguistic: "at the phonological and grammatical level they are so close that they appear to be one language" (Schmidt, 1999: xiv), and the differences between Hindi, Urdu and Punjabi lie in the area of morpho-phonology (Bhatia, 1993). There are no crucial differences between these languages with respect to the expression of definiteness. Hindi (Kachru, 2006), Urdu (Schmidt, 1999), and Punjabi (Bhatia, 1993) do not have a category of article. Hindi uses other determiners to mark indefiniteness, for instance the numeral *ek* 'one' or the indefinite pronoun *koi* 'some, any', as in *koi lerka* 'some boy', *ek mehila* 'one lady'. Similarly, Punjabi (Bhatia, 1993: 218) can use determiners such as the numeral *ikk* 'one' or pronouns *koi:/kujj* 'some, any', and demonstrative adjectives *e* 'this/these' and *o* 'that/those' to express the semantic contrast of definiteness/indefiniteness in combination with the features of proximity or remoteness.

In addition, word order can be used to mark definite nouns. Compare the two examples below:

(10) a. mez pər **kitab** hɛ
table on book is
'There is **a book** on the table'

b. **kitab** mez pər hɛ
book table on is
'**The book** is on the table'

Generally, the marking of the contrast definiteness/indefiniteness is not grammaticalized in Hindi, Urdu, and Punjabi and is not obligatory, and potential

ambiguity is resolved using pragmatic factors. Thus, we assume that these languages are similar to Chinese, Korean and Japanese in that they have not grammaticalized the category of definiteness.

3.3.5 Comparison across L1 backgrounds

Although all languages presumably have the semantic concepts underlying the features definiteness and specificity (they are semantic universals according to Ionin, 2003), languages do not select the same semantic contrasts for mapping onto morphosyntactic forms. It can be assumed for the purposes of this study that English, Spanish, and Arabic grammaticalize the feature [definite], whereas languages like Mandarin/Cantonese Chinese and Hindi/Urdu/Punjabi do not. Languages like English, Spanish and Arabic have a grammaticalized expression of definiteness in the shape of special functional morphemes – articles. Even among languages with article systems there is variation, so Spanish has a contrast between indefinite and definite articles, and Arabic has a contrast between a zero article and a definite article. Thus, the division of L1s in this study (Spanish versus Arabic versus Mandarin/Cantonese Chinese and Hindi/Urdu/Punjabi) is particularly interesting with respect to the investigation into the role of L1 in L2 acquisition.

3.4 The auxiliary verb system of English

Auxiliary verb learning in child L2 acquisition is of particular interest for several reasons. English auxiliary verbs are a part of most well-formed declarative, interrogative, and negative sentences. For instance, in order to produce correct declarative sentences such as *The elephant is running* or *The elephant was running*, it is necessary to use the appropriate auxiliary form *is* or *was* and to know that the auxiliary verb must be followed by the present participle form of the main verb *run*. It is also necessary to know the appropriate semantic contexts for the continuous forms *is running* or *was running*, as opposed to, for instance, *has run* or *ran*. Furthermore, in order to form a yes/no question or a *wh*-question, it is necessary to reverse the order of the auxiliary and the subject: *Is the elephant*

running? or *Why is the elephant running?* It is especially interesting to consider the acquisition of auxiliaries in structural contexts, rather than study the correct use of isolated forms, since auxiliary use is a wide-ranging task involving syntactic, morphological, and semantic knowledge. Auxiliary verbs are crucial for the structure of learner utterances from the very onset of acquisition because auxiliaries are very frequent and used in a variety of constructions. However, these verbs have low semantic content in isolation because their main function is to express tense, aspect and mood of the main verb.

There is a fair amount of evidence in the literature about patterns of auxiliary acquisition in monolingual children acquiring English as an L1, but not in child L2. Thus, there is comparative value in the investigation of child L2 acquisition of auxiliaries, because it can illuminate the similarities and dissimilarities between L1 and child L2. Finally, there is an important theoretical discussion in the literature concerning productivity in the early speech of L1 as well as L2 learners, and the role of L1 in child L2 acquisition. Therefore, research on the auxiliary system provides an opportunity to evaluate the predictions of the most widely known theories of language acquisition, the generativist theory and the constructionist theory, discussed in Chapter 2.

3.4.1 A generativist analysis of auxiliaries

From a generativist point of view, aspectual auxiliaries (BE and HAVE in English) can be formally analyzed as subcategorizing an Aspect Phrase (AspP) which contains functional features corresponding to perfective and progressive aspect (e.g. [+/-perfective], [+/- progressive]) (Adger, 2003). The functional projection AspP is placed between the Inflectional projection and the verb phrase. The formal account of the structure of declarative sentences can be applied to other languages as well, with Inflection and Aspect phrases being projected in the sentence structure (e.g. see Zagona (2002) for Spanish). For languages with subject-verb inversion in questions, it was argued that the word order can be accounted for in terms of the movement of the inflected verb from its position in

the Inflection Phrase (IP) to the Complementizer Phrase (CP). However, overt verb movement to CP in questions does not exist in all languages.

The acquisition of syntactic inversion in questions needs to be considered as an issue in its own right. Numerous studies have been dedicated to the acquisition of the subject-auxiliary inversion in L1, and there is still no agreement regarding what constitutes the mastery of word order in questions and how it is acquired (see, e.g. Theakston & Rowland, 2009a,b). According to generative approaches (e.g. Santelmann, Berk, Austin, Somarshekar & Lust, 2002), the syntactic operation of inversion is a universal option available from UG, but this option is not employed in all languages. Extending this idea to L2 acquisition together with the view that all UG options are acquirable in L2, we can expect L2 children to acquire the operation of inversion successfully. In addition, if we consider the possibility of transfer of L1 word order, learners whose L1s employ subject-auxiliary inversion are expected to have an advantage at the initial stage of acquisition. However, different predictions need to be made for the acquisition of *do*-support in questions, since the formation of questions with DO auxiliary is a structure that is specific to English. The mechanics of DO-question formation are represented as insertion of DO into I node in the structure followed by movement of DO from I to C (e.g. Adger, 2003; Santelmann et al., 2002). Since L1 transfer cannot facilitate L2 acquisition of *do*-support, all L2 children are expected to learn the DO-insertion and movement of auxiliary from input.

3.4.2 Auxiliaries in the constructionist approach

Constructions cut across the different levels of grammatical analysis, there are no abstract functional categories such as Tense or Aspect and as a consequence there is no discrete cutoff point between lexicon and grammar. The constructionist approach does acknowledge the existence of grammatical categories such as noun, verb, or auxiliary, but they are thought to exist only to the extent that they can be derived from the distributional regularities of the input. Thus, from the point of view of constructionist approach, auxiliaries are items that form part of constructions. Recall from the discussion of this approach in Chapter 2 (Section

2.3) that one utterance can be analyzed on different levels of abstractness. Consider the example from Goldberg (2006: 10). The utterance in (11) involves all of the constructions listed in (12):

(11) What did Liza buy Zach?

(12) a. *Liza*, *buy*, *Zach*, *what*, *do* constructions
b. Ditransitive construction
c. Question construction
d. Subject-auxiliary inversion construction
e. VP construction
f. NP construction

While the structure of the utterance includes a number of abstract, i.e. lexically unspecified, constructions, no underlying syntactic structures or phonologically empty elements are posited. Instead, it is argued that some constructions do not specify word order, while some constructions do. For instance, the ditransitive construction does not specify the order of the NPs and the VP, but the question construction does. In sum, the constructionist theory adopts a "what you see is what you get" approach to syntactic form.

Recall that in this view, children begin the language-learning process by acquiring a small number of lexically specific combinations that appear in the speech they hear. With respect to auxiliary suppliance in particular, children are expected to begin with a series of independent constructions, for example, *I V-ing* and *he's V-ing*. Such early constructions differ from adult-like constructions because they can contain low-scope slots (e.g. with only three verbs occurring in the *V-ing* slot) or they can be incorrect (e.g. *I V-ing*). Over the course of development, children are assumed to induce increasingly abstract constructions by developing variable slots in previously lexically specific frames. Thus they might only gradually develop a more abstract construction *N BE V-ing*. Children's early use of memorized utterances such as *I don't know* or *What's that?* points to another important frequency effect: not only the frequency of individual words

affects their acquisition, but also the input frequency of the words that precede or follow them (their collocates). As pointed out by Bybee (2007: 17), in order for frequency counts to be useful, it is often a more specific and longer string that must be counted rather than a more general and shorter one. The example cited by Bybee is the reduction of *don't*: it is the frequency of phrases such as *I don't know*, *I don't think*, *I don't have* that leads to the most extreme reduction of *don't* rather than the frequency of *don't* itself. According to Bybee, the effect of frequency is so strong that even in adult language some lexical phrases such as *I don't know* may be stored as linguistic wholes due to their high token frequency and coexist with more abstract constructions such as N DO+*not* V.

3.4.3 Previous acquisition studies of auxiliaries

Auxiliaries in L1 acquisition

While the present study is concerned with L2 acquisition, there is a growing body of work on the acquisition of auxiliaries in L1 acquisition, whose findings will be relevant to the issues investigated in this study.

There seems to be consensus regarding overall characteristics of L1 acquisition of auxiliaries (Klima & Bellugi, 1966; Pinker, 1984; Richards, 1990; Valian, 1991). The earliest combinations involving auxiliaries appear with only one main verb or in only one form (e.g. *don't* but not *doesn't*). Once children start producing auxiliaries, there is a long period in which the auxiliary forms that the child can produce are also frequently omitted (e.g. the child can alternate between saying *I'm playing* and *I playing*). Finally, most studies report very few errors of commission (i.e. structures such as *I is playing*). The point of disagreement lies in what precisely is postulated as linguistic knowledge of the child, whether it is innate or not, and how it is reflected in the use and nonuse of auxiliaries. It was mentioned in Chapter 2 (Section 2.3) that researchers working in the emergentist framework see linguistic knowledge as an abstraction from the actual use of language. In the generative framework, children are credited with a pre-given knowledge of abstract categories, and their task in acquisition is to map the words of their target language onto pre-existing abstract categories.

A growing number of researchers provide evidence for the role that input frequency plays in children's language development and for the emergentist view of L1 acquisition. It has been found that forms (both individual auxiliaries and multiword combinations, such as subject-auxiliary) that are highly frequent in the input are among the earliest to appear in child speech. For instance, researchers that have investigated the acquisition of declaratives and yes/no- and *wh*-questions reported that patterns of acquisition can be related to the language children hear (Rowland, 2007; Rowland & Pine 2000; Rowland et al. 2005; Theakston et al. 2005). One of the first studies that systematically analyzed the item-based nature of auxiliary use in L1 was Richards (1990). In his longitudinal study, Richards investigated the acquisition of auxiliaries BE, DO, and HAVE and modal verbs in a group of 30 children, looking at syntactic constructions with auxiliaries and also at the functions associated with the class of auxiliaries, namely negation, inversion, ellipsis, and emphasis. Negation of the auxiliary is accomplished by adding the particle *not* or *n't* (e.g. *did not*, *didn't*). Inversion of the subject and auxiliary is found in yes/no questions and most *wh*-questions (e.g. *Am I…?*, *Is it…?*). In ellipsis, the main verb is dropped and the auxiliary remains as the operator marking tense and agreement (e.g. *You know him but I don't*), and emphasis is the use of a stressed auxiliary form to mark contrastive emphasis or contradiction (e.g. *He does know you*). Richards found that after 9 months, less than half of the children had produced tokens of all the four functions. Rapid development in one part of the system (e.g. negation), contrasted with piecemeal development in other parts (e.g. inversion). The auxiliary class seemed to be well established in declarative utterances after 9 months, but children manifested considerable variation in the range of forms that they used and in the overlap between the verbs used with these different forms. Richards concluded that the children were slow to achieve any generality across the auxiliary system and that they developed a particular class of operators for each of the four functions rather than an auxiliary class as a whole.

A series of more recent studies specifically aimed at the acquisition of auxiliary BE suggested that different forms of BE are acquired separately and the

relation between constructions develops gradually (Rowland et al. 2005; Theakston & Lieven 2008; Theakston & Rowland, 2009a,b; Theakston et al., 2005). A detailed analysis of children's utterances with auxiliaries revealed that children tended to produce more correct utterances with higher frequency forms of BE and that they substituted lower frequency forms with higher frequency forms. For instance, children produced more correct responses with the higher frequency form *is* and substituted *is* for the lower frequency form *are*. Another factor that might be expected to affect children's auxiliary use and nonuse is the frequency of individual subject–auxiliary or *wh*-word–auxiliary combinations present in the language children hear. A number of studies demonstrated that the frequency of lexically specified constructions organized around *wh*-words, auxiliaries, and subjects predicted children's use of correct questions (Dabrowska, 2000; Dabrowska & Lieven, 2005; Rowland, 2007; Rowland & Pine, 2000).

Auxiliaries in child L2 acquisition

English verb morphology is one of the areas that have been substantially researched in child L2 acquisition, but most of previous studies were not focused on auxiliaries in particular. Early studies of L2 morphosyntax (Dulay & Burt, 1973, 1974) found that the order of acquisition of morphemes in child L2 was similar to the order reported in child L1 (Brown, 1973). Dulay & Burt's findings indicated that free morphemes such as auxiliary BE were acquired earlier than bound morphemes such as third person singular *–s* or past tense *–ed*. More recent studies provided support for this finding (Haznedar, 2001; Ionin & Wexler, 2002; Paradis, 2005). Furthermore, available reports suggest that acquisition rates differ for each auxiliary and that children are unlikely to acquire all auxiliaries in parallel. For instance, Jia & Fuse (2007) and Paradis (2005) demonstrated that BE is acquired earlier than DO. These disassociations distinguish L2 children from children acquiring L1 English (Ionin & Wexler, 2002; Rice, Wexler, & Hershberger, 1998), in which the contrast between free and bound tense-related morphemes is not so noticeable. Another difference between child L2 learners and L1 learners is the number of omission and commission errors. While omission is

the common type of error in all learners, larger proportions of commission errors were documented in child L2 learners than in L1 learners (Paradis, 2005). There are also a number of similarities between L1 and child L2 acquisition, namely the difficulty that all children have with acquiring the question structure, which results in higher error rates in questions. This difference was documented in structures with the auxiliary BE (Gavruseva, 2002, 2008; Ionin & Wexler, 2002; Paradis, 2005; Paradis et al., 2008) and with auxiliary DO. While some studies reported fast acquisition of DO in negative declarative structures, studies of DO in questions found that this structure was acquired later than both declaratives and questions with BE (Paradis, 2005; Paradis, Rice, Crago & Marquis, 2008).

As for the acquisition of the auxiliary DO, two case studies of children acquiring L2 English (Hakuta, 1974 and Ravem, 1978) did not find L1-based errors in DO contexts. In a recent case study, Haznedar (2007) reported that DO structures appeared early (after five months of exposure to English) and the use of *do*-support was consistent afterwards. However, all these studies investigated the use of DO in spontaneous speech, and Haznedar in particular pooled together the obligatory contexts for *do*-support in declaratives with negation and in questions. Bearing in mind that existing studies of the use of DO in questions report high rates of error, and that *do+not* negation is known to be acquired early in L1 acquisition, it might be the case that the use of DO in spontaneous speech created a misleading impression of the children's accuracy.

It also needs to be emphasized that there are no studies systematically analyzing the acquisition of auxiliaries BE and DO in child L2 acquisition, with the exception of Gavruseva and Ionin who focused on the auxiliary BE. Most child L2 studies investigated the general area of morphosyntax (Lakshmanan, 1995; Ionin & Wexler, 2002; Haznedar, 2001; Paradis, 2005) or tense and aspect (Gavruseva, 2008; Haznedar, 2007; Ionin, 2008), and the acquisition of auxiliary verbs per se was not the goal of this research. Thus, there has been little discussion of how the syntactic aspect of auxiliary verbs develops. However, bringing together the findings in previous work on child L2, we can see several trends, namely that omissions of BE are frequent in child L2 acquisition, that the

sentence structure plays a role, as declaratives with BE are acquired earlier than questions, and that questions with DO appear to be the most difficult structure. Finally, auxiliary use in spontaneous speech needs to be interpreted with caution, as it may not be a good representation of mastery.

3.5 Auxiliary systems in languages represented in the study

3.5.1 Arabic

Present tense of the verb in Arabic is used to refer to incomplete, ongoing actions or ongoing states. It corresponds to both the English present and present continuous tenses (Ryding, 2005: 442):

(13) a. a-ktub-u
'I write' or 'I am writing'

b. na-drus-u
'We study' or 'We are studying'

The present tense is formed from the present tense stem of the verb with the addition of prefixes and suffixes marking person, number, and mood. Thus, unlike English, Arabic does not have an auxiliary verb construction to convey the idea of continuous action. In the past tense, however, Arabic has a compound verb form denoting continuous action, somewhat similar to the English past continuous form *was*/*were* + V-*ing*. The construction consists of the verb *kanna* 'be.PAST' and a present tense form of the main verb. Unlike in English, both the auxiliary and the main verb are inflected for person, gender, and number. Other differences with English include word order (if there is an overt subject, it comes between the auxiliary and the main verb) and semantic ambiguity (the Arabic past tense construction can also convey habitual meaning similar to the English 'used to + V' construction) (Ryding: 446-447).

(14) kaanat ummi tatbuxu alghadaa
be.PERF.3SG my.mother cook.IMPF.3SG the.lunch
'My mother used to cook/was cooking (the) lunch'

Yes/no questions are formed by changing the intonation of the declarative sentence or by adding an interrogative prefix (*hal* or *ʻa*) to the first word of the declarative sentence (Ryding: 406).

(15) **ʻa**-haadhaa samiir-un?
prefix-this Samir
'Is this Samir?'

Unlike in English, there is no shift in word order to signal yes/no question formation.

3.5.2 Spanish

The progressive aspect in Spanish is marked using an appropriate form of *estar* 'to be' and the gerund form of a verb: *María está cantando* 'Maria is singing', *estoy hablando* '(I) am talking' (Butt & Benjamin, 1994; Zagona, 2002). The construction is syntactically very similar to the English progressive construction. However, according to Butt & Benjamin (1994: 231), "the Spanish continuous adds a nuance to, but does not substantially alter the meaning of the non-continuous verb form, so that the two forms are sometimes virtually interchangeable":

(16) a. Ana lee
Ana read-3SG
'Ana is reading'

b. Ana está leyendo
Ana be.3SG reading
'Ana is reading'

In yes/no question formation, the auxiliary verb *estar* 'be' moves to the sentence-initial position, similarly to English. Unlike in English, however, the word order

of the gerund and the inverted subject is flexible: the subject can almost always be placed immediately after the finite verb or immediately after the non-finite verb.

(17) ¿Está María cantando? Or ¿Está cantando María?
'Is Maria singing?'

Unlike in English, SV order is commonly used in yes-no questions, in which case only the rising intonation shows that a question is intended (Butt & Benjamin: 467).

(18) ¿Mamá ha comprado leche? (or ¿Ha comprado mamá leche?)
'Has mother bought any milk?'

However, in such sentences VS order is preferred if there is no object: *Ha llamado mama?* 'Has mother called?'.

3.5.3 Mandarin and Cantonese Chinese

These languages have no auxiliary verbs similar to the English *be*. Progressive aspect is marked with particles. In Mandarin, the counterpart of the progressive construction in English is the post-verbal particle *zhe* and the pre-verbal adverb *zhengzai* (or simply *zheng* or *zai*). The adverb and the particle can be used separately or simultaneously (Lin, 2001: 174).

(19) a. tamen **zhengzai** chi fan
they PROG eat meal
'They are eating their meal'

b. tamen chi **zhe** fan ne
they eat PROG meal NE(sentential particle)
'They are eating their meal'

In Cantonese, (Matthews & Yip, 1994: 200), the aspect markers are bound forms, behaving essentially as suffixes: in their functions as aspect markers, they may not be separated from the verb.

(20) *wohng siuje gong-**gan** dihnwa*
Wong Miss talk-PROG telephone
'Miss Wong is on the phone' (Matthews & Yip, 1994: 198)

In contrast with English, Spanish and Arabic, Mandarin marks tense lexically using time adverbs such as *zuotian* 'yesterday', *jintian* 'today', and *gangcai* 'just now' (Lin, 2001: 168). In formal syntactic analyses of tense, it has been argued that there is no syntactic distinction of finiteness/non-finiteness in Chinese languages, and thus the projection of Tense (or Inflection) is absent (Hu, Pan & Xu, 2001; Leung, 2003). Yang (1995) argues for Mandarin that aspect markers exist in the lexicon as intrinsic features of lexical verbs. Aspect markers are different from inflection markers because when a verb is selected from the lexicon, it already contains a set of aspectual features morphologically instantiated by *–zhe*, *–le*, or *–guo*, similarly to the English *-ing*. Yang argues that since aspectual features are grammaticalized as verb affixes, they need to be accounted for in terms of a functional category Aspect in the sentence structure. Under such an account, functional projections in clauses in Mandarin (and presumably also in Cantonese) would include AspP and VP, but no IP.

Question formation in Mandarin and Cantonese does not require the change of order of the subject and verb. Yes/no questions are formed by adding an interrogative sentential particle at the end of a normal declarative sentence: the particle *ma* in Mandarin and *maa* in Cantonese.

3.5.4 Hindi, Urdu and Punjabi

In Hindi, progressive aspect is marked with the auxiliary *reha*[3] plus the root of the main verb. Tense is marked by a separate auxiliary *hona* 'be' that follows the aspect auxiliary. Thus, unlike in English, present and past progressive forms are marked with a combination of two auxiliaries rather than one (Kachru, 2006: 148-150):

(21) kʰa rəha hɛ
eat PROG.SG PRES.SG
'He is eating'

(22) ja rəha tʰa
go PROG.SG PAST.SG
'he was going'

Present tense conveying the meaning of a habitual action is also expressed with an auxiliary verb construction that consists of an imperfective form of the main verb followed by the tense auxiliary *hona* 'be':

(23) a. cəl-ta hū
move.IMPF.SG PRES.1SG
'I move/go'

b. bol-ta hɛ
speak.IMPF.SG PRES.3SG
'He speaks'

Urdu (Schmidt, 1999) and Punjabi (Bhatia, 1993) tense and aspect systems are very similar to that of Hindi, with minor morphophonological differences in the forms of the auxiliary verbs or verb endings. For instance, compare Hindi third person singular form of the present tense auxiliary *hɛ* and Urdu *hai*, Hindi 3sg past tense auxiliary *tʰa* and Urdu *tʰā* (Schmidt 1999: 87-95).

[3] The auxiliary *reha* is homophonous with the perfect participle form of the verb *reh* 'live, stay, remain'

Word order in questions in Hindi, Urdu and Punjabi is the same as in declarative sentences, with interrogative pronouns replacing the original constituent. Consider the following examples from Hindi (Kachru, 2006: 186):

(24) **kɔn** mumbəi ja rəha hɛ?
who Mumbai go PROG.SG PRES.SG
'Who is going to Mumbai?'

(25) sərita ne us dukan se **kya** xərida?
Sarita CASE that shop from what buy.PERF
'What did Sarita buy from that shop?'

Spanish, Arabic, and Hindi are similar in that they grammaticalize the features [perfective] and [progressive]. Applying the formal syntactic account to these languages, the structure of declarative sentences can be described as having Inflection and Aspect phrases projected. However, in Arabic and Hindi overt verb movement to Complementizer Phrase, the top layer of syntactic structure, does not take place in questions.

To summarize, Arabic, Hindi, and Spanish have a full range of progressive constructions with an auxiliary similar to the English BE inflected for person and number. In Arabic, however, the progressive construction is used only in the past tense. Only Spanish has subject-verb inversion in questions (BE and main Verbs). None of the languages have an auxiliary similar to DO in yes/no questions. As for word order in questions, Spanish is the only language with subject-verb inversion in questions. However, in Spanish the inversion rule applies both to auxiliaries and to main verbs, whereas in English it applies only to auxiliaries. Furthermore, questions with uninverted subject-verb order appear to be more acceptable in Spanish than in English (Butt & Benjamin, 1994), since in English the uninverted word order is typically used to mark surprise or request for clarification rather than a neutral question. A corpus study is needed to estimate the frequency of uninverted questions in the two languages, which is beyond the scope of my study. What is important to know for the formal syntactic analysis and how it

translates into the L1 transfer models is that Spanish is the only language that employs the syntactic operation of subject-verb inversion in questions.

3.6 Predictions for child L2 acquisition of articles and auxiliaries

General theoretical predictions for language development in child L2 learners were discussed in Chapter 2 (Sections 2.2 and 2.3). The sections that follow specify the predictions of the generative and constructionist approaches with regard to articles and auxiliaries.

3.6.1 Full Transfer / Full Access

Articles

Researchers working on acquisition in the generative framework argued that the presence of articles in a language signals the presence of the abstract category of Determiner. If functional projections are initially transferred from L1, DP structure is also expected to be transferred. Following FT/FA, speakers of languages like Spanish or Arabic are expected to transfer the knowledge of the category Determiner from their L1 to English. If a learner's L1 does not have the D category, then that learner will have to establish the presence of this category in L2 from input and add a functional layer to their noun phrases. According to this account, learners from such backgrounds are expected to omit articles at the onset of L2 acquisition, because their initial L2 functional structure (coming from their L1) might not include a D projection, or the necessary semantic features assigned to lexical items like articles. It follows from this model that learners from [+article] L1 backgrounds will be more accurate with articles at the early stages of L2 acquisition.

Auxiliaries

With respect to the auxiliary verb system, under the generative view auxiliaries are understood as the lexical items that are the spell-out of the formal features of tense and agreement. FT/FA model would predict that the knowledge of abstract grammatical categories such as Tense and Aspect is transferred from a

L2 learner's L1. If a certain category is not instantiated in L1, it will be supplied from the universal inventory of categories that learners have access to. But at the early stage, errors are expected in learners whose L1s do not instantiate categories of Tense and Aspect. Children from Chinese L1 background are expected to be less accurate with auxiliaries than all other groups.

(26) Full Transfer/Full Access predictions

a. Children whose L1s project Tense and Aspect in the sentence structure are expected to transfer this knowledge to L2 English and be more accurate with the provision of BE than the children whose L1s do not instantiate these categories.

b. All children are expected to have difficulty with DO-questions because it is an idiosyncratic structure of English, resulting in lower accuracy with DO than with BE

3.6.2 Feature assembly

Articles

We also need to ask whether there are differences to be expected between groups of learners whose L1 article systems are parallel to English (i.e. they have a definite and an indefinite article) and those whose L1s are not completely similar (e.g. Arabic has a definite article but no indefinite article). Such differences cannot be captured by the FT/FA model in its original form since it only analyzes L1 transfer as transfer of abstract categories such as D, rather than specific lexical items. Cross-linguistic differences such as the one between Arabic and English are better captured by a feature-assembly approach (Lardiere, 2009).

Applying a feature-assembly approach to article systems, we can say that articles are exponents (phonological representations) of the functional category D, which can host formal features such as [±definite], [±specific], or [±singular]. Featural contexts of insertion for articles are as follows (Hawkins et al. 2006: 20): *a* is inserted in [-definite, +singular] contexts, *the* is inserted in [+definite] contexts. Since there is no article for indefinite plural nouns in English, the

featural context for such nouns is [-definite, -singular]. Thus, in order to use *a* appropriately, learners have to identify not only the feature [definite], but also the feature [singular] as relevant for the insertion of this article.

Under a feature-assembly account, L2 learners start off with a fully assembled set of L1 lexical items and grammatical categories. Wakabayashi (2009: 339) hypothesized that at the early stages of L2 acquisition learners assemble a new feature only to a prototypical lexical item and often fail to include the feature into numeration altogether. Applying this approach to articles, we expect the L2 learners who have no articles in their L1 to omit articles in L2 English or to use only one prototypical (featurally simplest) form as the substitution form. Which article is expected to be the prototypical form? In a discussion of article acquisition in an adult L2 learner, Lardiere (2004: 335) argued that "definite articles in English need not take number and the count/mass distinction into account, which makes them less featurally complex than indefinites in at least one respect". Thus, the definite article is likely to be used as the substitution form.

This analysis is promising for the comparison of learners with L1 backgrounds that differ in the particular manifestation of the article system, for instance languages with no articles (Mandarin and Cantonese) compared with languages that have only one article (Arabic), in turn compared to languages with a two-article contrast (Spanish). These differences are summarized in Table 3.1 below.

Table 3.1

The article system: Feature mapping in English and children's L1s

	Formal features	Mapping onto surface forms	
	[±definite]	Definite article	Indefinite article
English	✔	✔	✔
Arabic	✔	✔	✗
Chinese	✗	✗	✗
Hindi	✗	✗	✗
Spanish	✔	✔	✔

Arabic speakers are expected to transfer the mapping of the [+definite] feature onto the definite article morpheme and the mapping of the [-definite] feature onto the null morpheme. In other words, they are expected to have difficulty with the indefinite article in English. Spanish-speaking children are expected to outperform all other L1 groups due to transfer of [±definite] feature mappings onto the definite and indefinite article.

Auxiliaries

Spanish-, Arabic-, and Hindi-speaking learners are expected to transfer the knowledge of the functional category Tense and Agreement to their L2 English. The difference among the groups is that the Spanish speakers tense- and aspect-related features are mapped onto vocabulary items (auxiliary verbs) in the same way as in English for progressive constructions with BE. In contrast, in Arabic these features are mapped onto overt auxiliaries only in the past tense, whereas in Hindi Tense and Aspect are mapped onto two separate auxiliaries. All these languages are contrasted with Mandarin and Cantonese Chinese, in which the abstract category of Tense is not projected, and Aspect features are mapped onto sentential particles.

Table 3.2

The auxiliary system: Feature mapping in English and children's L1s

	Formal features		Mapping onto surface forms	
	[±past]	[±progressive]	Auxiliary for Tense and Aspect	Subject-auxiliary inversion
English	✔	✔	✔	✔
Arabic	✔	✔ *	✔ *	✗
Chinese	✗	✔ / ✗ **	✗	✗
Hindi	✔	✔	✔	✗
Spanish	✔	✔	✔	✔

* Only in the past tense

** Aspect markers are considered lexical rather than grammatical in some analyses (see Section 3.5.3)

With respect to learners' accuracy, the feature assembly approach allows us to make the predictions as follows:

(27) a. For auxiliary BE, accuracy in L1 groups is expected to be distributed as follows: Spanish > Arabic, Hindi > Chinese;
b. All four L1 groups are expected to demonstrate low accuracy with DO.

Furthermore, both the FT/FA and the feature assembly account predict L1 effects in the acquisition of syntactic inversion in questions: If the Spanish children transfer the knowledge of inversion from L1, there should be no difference between their accuracy with declaratives and questions with BE, because inversion of BE employs the same syntactic operation in English and Spanish. If the Arabic-, Chinese-, and Hindi- speaking children transfer the syntactic structures from their L1, we expect to find their knowledge of inversion to be initially faulty, since their first languages to not require any changes in word order for question formation.

3.6.3 The usage-based model

Articles

The role of syntactic constructions and input frequency in the distribution of definite and indefinite articles in L2 children will be investigated in terms of the usage-based model and the following predictions will be made:

1. If the frequency of articles in the input is reflected in the order of acquisition, *the* will be acquired before *a*;

2. The distribution of articles with respect to construction type in children's speech will be similar to that in the input:

- Article use will be influenced by collocates of articles, e.g. the preceding verb in the case of direct object NPs.

- The use of articles at the early stages of acquisition will be construction-specific, with gradual increase in the variety of contexts in which articles are used correctly

Auxiliaries

The trends in the growth of productivity in auxiliary constructions noted in the previous studies of L1 acquisition (Section 3.4.3) are equally applicable to child L2 acquisition. The following predictions can be formulated for auxiliary constructions:

1. Since each auxiliary is regarded as the range of specific inflected forms, rates of accurate suppliance of different forms of the same auxiliary are expected to differ depending on the input frequency of each form. The more frequent form will be acquired first.

2. The distribution of auxiliaries with respect to construction type in children's speech will be similar to that in the input:

- The type of subject is expected to influence accuracy with auxiliaries that follow it because subjects with high input frequency (e.g. pronouns) can be chunked together with auxiliaries at the early stages of acquisition (e.g. *I'm*) (Theakston et al. 2005; Theakston & Rowland, 2009a)

- In building constructions of increasing complexity, children are expected to start with constructions that are lexically specific and build up towards higher abstractness. Importantly, this should apply to constructions of all levels of complexity: sentence-level constructions with embedded clauses (Diessel & Tomasello, 2000; 2001), as well as word-level and phrase-level constructions.

CHAPTER 4. METHOD

This chapter presents detailed information on the methodology used to collect and score data from L2 children with the aim of exploring their use of article and auxiliary verb systems in English. Section 3.1 presents the details about the children that participated in the study and the selection and grouping of children for particular analyses. Section 3.2 describes the procedures of data collection and how the use of articles and auxiliaries was scored. Details about the analyses of the children's scores and the results will be discussed in the respective chapters on articles (Chapters 5 and 7) and auxiliaries (Chapters 6 and 8).

4.1 Participants

4.1.1 The longitudinal study

The research questions will be addressed by analyzing data from two sources, a longitudinal and a larger cross-sectional study. Participants in both studies were children learning L2 English in Edmonton. The longitudinal corpus included data collected from 24 children in the study of various aspects of English L2 development involving multiple measures of lexical, grammatical and narrative performance. The children's mean age was 5;4 and mean exposure to English was 9 months at the onset of the study. It is important to point out that for this study onset of exposure to English was defined at the child's entry into a daycare, preschool or school program conducted in English. Therefore, it is possible that children could have had some limited pre-exposure to English if they were born in Canada or arrived before entry into a program. However, in all households, the L1 was the primary or exclusive language used by parents to their children – this was an inclusion criterion for the study.

For the analyses presented in this thesis, several subsets of the group of 24 participants were selected. For the analysis of the children's use of articles in Chapter 5, data from 16 children will be used. Table 4.1 summarizes the information about this subset of participants.

Table 4.1

In the longitudinal study, children's L1, age at the onset of the study, and months of exposure (MOE) to English

Participant code	L1	Age	MOE
LLKC	Arabic	4;10	11
TRRK	Arabic	4;02	8
YSSF	Arabic	4;11	9
CHRS	Romanian	6;02	5
DVDC	Spanish	6;03	8
SBST	Spanish	5;01	15
SMNS	Spanish	5;07	6
RNL	Cantonese	4;08	16
RMLM	Japanese	4;02	9
GSYN	Korean	5;02	2
DNNS	Mand/Cantonese	4;07	7
CNDX	Mandarin	6;09	8
DNNC	Mandarin	5;09	9
MRSS	Mandarin	5;00	4
JHHN	Mandarin	5;11	18
TNYN	Mandarin	6;07	7
Mean		5;4	9

The objective of selecting a smaller group of participants was to exclude children whose data collection was interrupted and also obtain two groups of participants of roughly the same size such with typologically similar L1s. For the analysis of the children's acquisition of articles, seven children whose L1s were Arabic, Romanian, and Spanish were grouped together as the [+article] group. Nine children whose L1s are Japanese, Korean, and Mandarin or Cantonese composed the [–article] group. All the children were tested every six months for two years. There were five rounds of data collection, and in each round samples of spontaneous and elicited speech were obtained. Tables 4.2 summarizes the information about the participants at each stage of testing.

Table 4.2

In the longitudinal study, average ages and months of exposure (MOE) at each round of data collection

Round	1	2	3	4	5
Age	5;4	5;10	6;4	6;10	7;4
MOE	9	16	22	27	34

For the analysis of the children's language development from a usage-based perspective, this group was further limited to include only 9 participants in order to analyze each child's data as a case study. These children were LLKC, TRRK, YSSF (L1 Arabic); DVDC, SBST, SMNS (L1 Spanish); and DNNC, MRSS, JNNH (L1 Mandarin). These children were chosen in order to have three L1 groups: Arabic, Spanish, and Chinese. From the Arabic and the Spanish groups, all the three children were included. From the Chinese group, three children were chosen so as to match the other two groups.

4.1.2 The cross-sectional study

The cross-sectional study will draw on data from the corpus of over 200 children 5;0-6;11 years old. For the analysis of the children's use of articles, 40 children will be selected. For the analysis of their use of auxiliaries, 48 children were selected. Only the children whose exposure to English was within the range of 2-18 months were included in the analysis. Since one of the goals of the study was to investigate L1 influence in child L2 acquisition, the study focused on the learners at the earliest stage of L2 acquisition when the effects of L1 are most pronounced. Children were divided into four groups according to their L1: Mandarin/Cantonese Chinese, Hindi/Urdu/Punjabi[4], Spanish, and Arabic. Tables 4.3 and 4.4 summarize the background information for the four groups of participants.

[4] In tables and figures, speakers Mandarin and Cantonese Chinese will be labeled as 'Chinese', and speakers of Hindi/Udru/Punjabi will be labeled as 'Hindi'

Table 4.3

For the analysis of articles, participants' L1, age in years; months, and months of exposure to English (MOE)

L1	# of participants	Age*	MOE*
Arabic	10	5;09 (4;10-6;08) 0;7	11 (2-18) 6.8
Chinese	10	5;11 (5;03-6;09) 0;5	9 (5-17) 3.8
Spanish	10	5;09 (5;00-7;00) 0;10	10 (4-18) 6.5
Hindi	10	5;08 (5;01-7;00) 0;6	9 (4-17) 4.4

*The numbers are given in the following format: mean (range) standard deviation

Table 4.4

For the analysis of auxiliaries, participants' L1, age in years; months, and months of exposure to English (MOE)

L1	# of participants	Age*	MOE*
Arabic	12	5;08 (4;10-7;00)	13 (2-18)
Chinese	12	5;09 (4;11-6;11)	14 (6-21)
Spanish	12	5;08 (5;00-7;02)	13 (4-21)
Hindi	12	5;09 (5;00-7;00)	12 (6-19)

Originally the 'early stage' group was limited to children with 18 or less MOE, but it resulted in very small groups for each L1 and in insufficient scorable responses, as the elicitation task was difficult for the children at the earliest (<6 MOE) stage of acquisition. It was necessary to widen the range of MOE in order to add more participants into the groups.

4.2 Procedures

Testing procedures in the longitudinal and in the cross-sectional study were similar. Data collection included recordings of a spontaneous speech sample, as well as English standardized tests of phonological memory, knowledge of vocabulary, accuracy with verbal morphology and narrative skills. Interviews and tests in both studies are videotaped and transcribed by research assistants who are native speakers of English. I will be using a subset of these measures.

4.2.1 Spontaneous speech

Spontaneous speech samples were obtained during play sessions with the research assistants, who were native speakers of Canadian English. Play sessions were semi-structured interviews, in which children were asked questions about their past and present activities, encouraging them to use a variety of morphosyntactic contexts. Play sessions were approximately 45 minutes long in the longitudinal study and 15 minutes long in the cross-sectional study. All sessions were videotaped and transcribed in CHAT format by the assistants. In order to ensure the accuracy of transcription, a reliability check was performed. Twenty percent of the recordings were re-transcribed by a second research assistant and reliability scores were calculated. The inter-transcriber reliability was 95%, and the mismatches in transcription were discussed and settled by the two transcribers.

Later CHAT files were manually coded for the use of several target morphemes, including articles, third person singular suffix *–s*, plural suffix *–s*, and auxiliaries. For the purposes of this thesis, only utterances with auxiliaries BE and DO were included in the analysis of spontaneous speech.

The majority of previous work on L2 children was based on data from spontaneous speech samples. However, relying only on spontaneous speech has some disadvantages. Utterances produced in freeplay situations can be an unbalanced representation of the children's mastery of the English morphosyntax. For instance, most of the children's utterances in spontaneous conversations in the present study had first and second person singular subjects. Furthermore, since in most cases the experimenters had to ask questions to encourage children to talk, children produced very few questions themselves. To compensate for these limitations, data from Test of Early Grammatical Impairment (TEGI, Rice & Wexler, 2001) were used to complement the data from the spontaneous speech samples. The test allowed us to evaluate children's use of auxiliaries *is* and *are* in declaratives and questions, and auxiliaries *do* and *does* in yes/no questions, which are naturally very infrequent in children's spontaneous conversations.

In the end, each child was assigned a percent correct, percent omission, and percent commission score for each auxiliary and article context. Each auxiliary was coded in isolation for correct use or error type. In addition, the structural context of auxiliary use was coded, namely the type of utterance (declarative or interrogative), type of subject (pronominal or nominal), and the main verb.

4.2.2 Elicited speech: Narratives

The children were visited in their homes or schools and participated in several tasks. The children were given a semi-structured interview aimed at spontaneous speech elicitation. Then, they were asked to tell stories based on picture books. All visits were videotaped and then transcribed in the CHAT format (MacWhinney, 2000). For the analysis of the children's use of articles, I analyze one part of the corpus, namely the part that includes transcriptions of narratives. Picture books used for the elicitation of narratives were designed as a part of the Edmonton Narrative Norms Instrument project (ENNI: Schneider, 2004a,b). The aim of the ENNI project was to develop a task for eliciting narratives in which characteristics of the stories were controlled. Such a task is similar to spontaneous speech recording because it encourages children to use their language skills to develop a story, but at the same time it makes children's stories comparable to one another.

The narrative elicitation materials involved two picture books, A and B, each containing a set of three stories. The picture stories were drawn by a professional artist on the basis of written scripts and then evaluated by a panel of specialists. In each book, the complexity and length of the stories increases from story 1 to story 3. Each picture sequence follows a typical story outline with an introduction part, an unexpected event or problem, a solution, and an outcome. Each sequence of three stories had two protagonists (two different animals), two secondary characters, and one or two objects that played a role in the story. The protagonists are the same throughout the three stories, and the secondary

characters and objects are added in the second and third stories, thus increasing the complexity of the cartoons. In addition to animate characters, in the first and the third story there are objects that play an important role in the story (a *ball*, an *airplane*, and a *net* in set A; and a *sandcastle*, a *balloon*, and a *bunch of balloons* in set B). The illustrations in Figure 4.1 represent the main parts of the first narrative out of six.

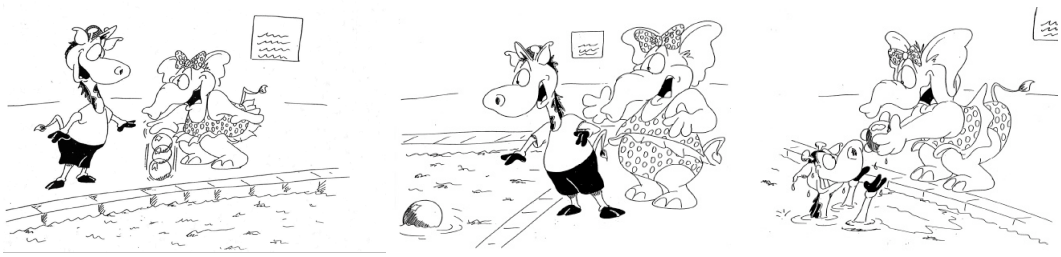

Figure 4.1. Story A1, Edmonton Narrative Norms Instrument

The advantage of narratives as a method of speech elicitation is that it allows us to tap the children's ability to present a situation that is unfamiliar to the hearer. An important part of this ability is the appropriate use of articles. In a study of the acquisition of the article system, the analysis of narratives can bring to light contexts of use that hardly occur in naturalistic speech that is mostly focused on the here-and-now. Since all the narratives were based on the same picture sets, we can partly control children's use of referring expressions and compare their stories. Narratives are useful in testing children's knowledge of articles because it is possible to clearly distinguish utterances where the new characters are introduced and the utterances where they are part of the common knowledge.

At the beginning of each session, the experimenter explained to the child that she did not know what the book was about and that she couldn't look at the pictures. When the children tried to point or use gestures to explain what the characters were doing, the experimenter reminded them that she could not see the pictures and that they had to use their words to tell the story. Before each story, the experimenter asked the child "Tell me your story" or "What is happening on the first page?". If the child hesitated because he or she didn't know the English

names for the animals in the pictures, the experimenter said, "You can use whatever name you want". Most importantly, the experimenter could not provide the names for the characters herself and did not mention the characters in her prompts because it could influence the child's use of articles.

Coding of narratives

In the process of scoring, only the use of articles was considered and all other ways to refer to characters, such as personal and deictic pronouns and proper names were excluded. Thus, the analysis was limited to the use of articles with singular countable nouns because these nouns required an article in all contexts. For the coding of nouns that met the inclusion criteria, two contexts were set apart, namely definite and indefinite contexts. The first mention of a referent in most cases was scored as an obligatory context for an indefinite noun, and all subsequent mentions of the same referent were scored as obligatory definite contexts. For each noun in indefinite and definite contexts, the use of the article was further scored as correct (*a* in indefinite contexts, *the* in definite contexts), substitution error (*the* in indefinite contexts, *a* in definite contexts), or omission error (null article, which was always incorrect, since proper names and plural nouns were not included in the analysis). There were some exceptions when the use of a definite noun was grammatical for first mentions, for instance when talking about the swimming pool and then introducing the character as ***the*** *lifeguard* (because pools always have lifeguards), or defining the noun with a relative clause: *Then* ***the*** *lady who cleans the pool showed up*. Such contexts were scored as definite, even though they were first mentions of new characters.

Consider the utterances from children's stories given in (1) and (2). In the first utterance in (1), indefinite articles are supplied correctly in an indefinite context. In the second utterance, *the* is used correctly with the definite nouns *elephant* and *donkey*, but it is used incorrectly with the indefinite noun *ball*, which is the first mention of this object in the story.

(1) There's **a elephant** and **a donkey**.
The elephant and **the donkey** throw **the ball** in the mud.
(028, L1 Arabic, 5;00, 18 MOE) [5]

(2) EXP: how do you start?
CHI: # mm # **the** elephant throw **the** ball.
(JNNH, L1 Mandarin, 5;11, 18 MOE)

In (3), we find an incorrect use of *the* with the nouns *girl rabbit* and *boy rabbit* that are mentioned in the story for the first same, but a correct use of *a* with the noun *castle*.

(3) **The** girl rabbit was making **a** castle, and then **the** boy rabbit came.
(088, L1 Punjabi, 5 years old, 17 MOE)

Example (4) illustrates *the* substitution in an indefinite context with the noun *balloon* and correct use of indefinite and definite articles with nouns referring to the characters *elephant* and *bunny*.

(4) Once upon a time, there was **a** elephant with **a** bunny.
The bunny see **the** balloon.
(019, L1 Spanish, 5;0, 6 MOE)

Omissions of articles are illustrated in (5) and (6). In the story given in (6), which clearly illustrates the child's inconsistency with article use, the same nouns *giraffe* and *elephant* are used intermittently with and without an article.

(5) **Giraffe** is um playing with his plane.
And he's, **the giraffe** is um flying it.
And then **the elephant** is um flying it.

[5] The child's name code, first language, age, and months of exposure to English (MOE).

And then **elephant** drop it in the water.
(040, L1 Urdu, age 5;7, 15 MOE)

(6) **A little elephant** in # in the pool.
…
Giraffe see it too.
That's not **a giraffe** [= whispered].
And **little elephant** got the ball.
(CNDX, L1 Mandarin, 6;09, 8 MOE)

Individual differences in children's storytelling

Certain peculiarities of narratives were taken into account in the coding procedure. For instance, some children used bare nouns as proper names to refer to the characters in the stories, such as *Doggie* or *Rabbit*. Such bare nouns were not coded as errors when the children were consistent in using them as names. Since the main characters in the three stories were the same, such nouns were considered proper names only when they were used without an article across all the stories. When children used both bare nouns and nouns with articles within the same story, as in (18) and (19) above, bare nouns were coded as errors. Such cases were different from the one illustrated in (20).

(7) story B1:
*CHI: like **doggie** was building s… sandcastle.
*CHI: then ra… **rabbit** came.
*CHI: then **rabbit** said "let's pour some sand on xxx".
*CHI: and **dog** said "maybe no".

story B2:
*CHI: like when **dog** was going to [/] for a picnic.
*CHI: he met **rabbit** on the # way.

story B3:

*CHI: like **doggie** he got a balloon.

*CHI: **bunny** said “can I hold it?”

(MRSS, L1 Mandarin, 7;00, 28 MOE)

In (7), we see that the child consistently used bare nouns to refer to the characters in all the three stories, therefore they were coded as names and not included in the ‘null article’ counts. In (8) below, I illustrate an inconsistent use of a bare noun, which alternated with the same noun with the indefinite article and the definite article within one story.

In data coding, it was also important to take into account the strategy that the children used in telling the story. Since the main characters were the same in the three stories in each picture book, some children told the three stories as parts of one continuous story. In this case, only the first mentions in the *first* story were coded as indefinite contexts. If some children started each of the three stories as an independent story, first mentions in *each* story were considered indefinite contexts.

(8) Story B1:

*CHI: once upon a time.

*CHI: there was **a** rabbit and two rabbits wanna make # a # castle.

Story B2:

*CHI: one upon a time.

*CHI: there's two rabbits # to get some berries.

Story B3:

*CHI: once upon a time.

*CHI: there was **a** rabbit and ano… two rabbits.

(YSSF, L1 Arabic, R4 6;05)

In cases like (8), each story was started with the introductory 'once upon a time', so I considered the first uses of referring expressions to be indefinite in each story. Then, the uses of *a* were coded as correct in each story.

To sum up the coding procedure, each noun included in the analysis represented one of the two kinds of obligatory contexts for articles, indefinite or definite, and there three possible codes for article use in each context:

Indefinite: - correct (*a*+noun, e.g. *He had a plane*)
- commission (*the*+noun, e.g. *He had the plane*)
- omission (bare noun, e.g. *He had plane*)

Definite: - correct (*the*+noun, e.g. *The plane fell into the pool*)
- commission (*a*+noun, e.g. *A plane fell into the pool*)
- omission (bare noun, e.g. *Plane fell into the pool*)

These coding categories are necessary to compare the L1 groups in two ways: first, to determine whether some L1 groups are more accurate than others (L1 facilitation effect), and second, to compare the types of errors that the children make (different L1s may result in more omission or commission errors). Commission errors can also indicate whether some forms are 'easier' to learn than others and are used as default substitutions, which is necessary for testing the feature assembly model.

Data from children who used less than three scorable definite or indefinite nouns were excluded from the counts.

4.2.3 Elicited speech: Test of Early Grammatical Impairment

In both the longitudinal and the cross-sectional studies, children's production of grammatical morphemes in elicited speech will be assessed using the grammatical probes of TEGI. This test was developed as a standardized measure for identifying children with SLI. It includes separate probes for third person singular, past tense, and the auxiliaries BE and DO. For the purpose of the present study, I will use the scores from the BE/DO probe. Auxiliaries BE and DO are

elicited in a play situation during which the experimenter encourages the children to ask questions to a puppet that can talk to toy animals, or make statements about the toys. For example, the experimenter says 'I wonder if the kitty is resting. Let's ask the puppet'. The target utterance for the child then will be 'Is the kitty resting?'. In order to elicit a statement with auxiliary BE, the experimenter would say, 'So the moon guys are jumping. Tell me about the bug'. In this case, the child's target response was 'The bug is jumping'. Importantly, the experimenters followed a strict scenario developed for the test and used only the prompts given in the manual. All the prompts were formulated in such a way that they did not include the target structure, thus excluding the possibility that the child might imitate the experimenter's utterances. For instance, in question elicitations, the experimenters did not produce any inverted questions themselves. In prompts eliciting a statement with auxiliary BE in the plural, the auxiliary was used in the singular, e.g. 'What's happening with the moon guys?'. Example (9) lists the three main structures elicited in the test.

(9) a. Prompt: *I wonder if the bug's resting. You ask the puppet about the bug*.
Target answer: *Is the bug resting?* (BE-question)

b. Prompt: *The moon guys are jumping*. Tell me about the bug.
Target answer*: The bug is jumping* (BE-declarative)

c. Prompt: You ask the puppet if the kitty likes hamburgers.
Target answer: *Does the kitty like hamburgers?* (DO-question)

This example provides illustrations for third person singular contexts. The test also included similar contexts for the elicitation of plural forms of auxiliaries BE and DO. In total, the BE/DO probe comprised 11 attempts of DO-questions, 6 attempts for BE-statements, and 6 attempts for BE-questions.

TEGI coding: Auxiliary BE

All obligatory contexts for the elicitation of auxiliary BE were third person (singular or plural) present tense. Utterances with no subject or with an auxiliary other than BE were excluded. Elliptical utterances that consisted of just the participle form of the main verb, as in (10), were excluded because such utterances could not be unambiguously judged as marked or not marked for finiteness, or it was not clear what the syntactic subject of the sentence was.

(10) Unscorable utterances:

a. EXP: So the kitty's taking a nap. What about the bears?
   CHI: Are resting.

b. EXP: What's going on with these bears?
   CHI: Having fun.

c. EXP: So the kitty's taking a nap. What about the bears?
   CHI: They are.

The responses that satisfied the inclusion criteria outlined in this section were further analyzed as follows. Uses of BE were coded as correct in obligatory contexts when the auxiliary was appropriately inflected for person, number, and tense and accompanied by the participle form of main verb. The errors in BE contexts were the omission errors (utterances with only the main verb present in participle form) and commission errors (utterances with a wrong form of BE). In questions with BE, errors also included double marking, when the question was formed with the auxiliary BE used twice, at the beginning of the sentence and after the subject. Examples of all types of scorable accurate responses and errors are provided in (3) for statements and in (4) for questions.

(11) BE in declaratives

correct: The bug is jumping.
They're having fun.
The bears are jumping too.

omission: The bug jumping.
The moon guys having fun too.
commission: The bug are jumping.

(12) BE in questions
correct: Are the moon guys crying?
Is the kitty resting?
omission: The moon guys crying?
The bears resting?
commission: Is the moon guys crying?
Is the bears feeling good now?
Are the bear crying?
double marking: Are the moon guys are crying?
Is the bug is laughing?

Thus, all scorable responses were full progressive constructions with the subject, auxiliary and participle present. In question elicitation, children sometimes produced questions with a correct form of BE, but in an uninverted declarative form, as illustrated in (13). Such utterances were coded as uninverted questions only when they were pronounced with what was clearly a rising question intonation.

(13) The kitty's laughing?

Such questions were not included for the accuracy counts, but they will be discussed separately in the analysis of error types.

TEGI coding: Auxiliary DO

Responses with DO were scored as correct when the auxiliary was used with correct tense and person agreement, and the main verb was uninflected.

Commission (wrong form) errors were uses of DO with wrong agreement (DO for DOES or DOES for DO), wrong tense (DID for DO), or wrong auxiliary (IS for DOES). Double marking errors refer to utterances with DOES in which the main verb was inappropriately inflected with 3sg *–s* and utterances with DID in which the main verb was inflected for past tense. In negative sentences, omission errors refer to sentences with the negation marker *no*/*not* and a main verb.

(14) DO in yes/no questions in TEGI
correct: Does the bug need a tissue?
commission: **Do** the bug need a tissue?
double marking: **Does** the bug need**s** a tissue?

Unlike in the coding of BE statements and questions, *omission* was not included as a type of error for DO-questions. In interrogative sentences, it was often difficult to tell whether the child omitted DO or used a non-inverted question with a finite main verb:

(15) a. The kitty needs tissue?
b. The moon guys like orange water?

For instance, if in a context that requires an auxiliary *does*, a child produced a sentence with a question intonation and a bare main verb it can be interpreted as the result of DO omission or as an uninverted question with 3sg *–s* omitted on the main verb:

(16) He want some juice?
= 'He want(s) some juice?' or
= '(Does) he want some juice?'

In other words, responses like these are ambiguous between the omission of DO and the omission of finite marking on the main verb. In the contexts for *does*,

children also produced utterances that had a question intonation and a main verb marked for 3sg. Such responses were not common, which could be partly due to children's inconsistent use of 3sg *–s* at this stage of acquisition. Thus, in the children's responses, there were three types of questions with the absence of auxiliary DO: questions with a bare verb in a third person singular context; questions with a bare verb in a third person plural context, or a verb inflected with third person singular *–s* in the absence of *do*-support.

(17) a. He want an apple?
b. They like juice?
c. He wants an apple?

All these responses were coded as *uninverted questions*. Since there were 11 contexts for DO in the elicitation task, I cannot analyze these error types separately because it will result in very few tokens per child. Therefore, all these types of non-target responses will be analyzed as absence of *do*-support.

The majority of previous work on L2 children was based on data from spontaneous speech samples. However, relying only on spontaneous speech has some disadvantages. Utterances produced in freeplay situations can be an unbalanced representation of the children's mastery of the English morphosyntax. For instance, most of the children's utterances in spontaneous conversations in the present study had first and second person singular subjects. Furthermore, since in most cases the experimenters had to ask questions to encourage children to talk, children produced very few questions themselves. To compensate for these limitations, data from TEGI were used to complement the data from the spontaneous speech samples. The test allowed us to evaluate children's use of auxiliaries *is* and *are* in declaratives and questions, and auxiliaries *do* and *does* in yes/no questions, which are naturally very infrequent in children's spontaneous conversations.

4.2.4 Adult speech corpus

The L2 input that the children received was mostly English spoken at school. Since the study was not aimed at documenting the English that the children in Edmonton heard from adults and other children, it was necessary to find a corpus of spoken English to use as a model of the speech that the children were exposed to. It was decided to use corpora of adult-to-adult speech rather than speech to children because most of the corpora available online contain speech to younger children (less than 6 years old) and also because most of the corpora contain parents' speech to their children. Since the children in the present study were older and received English input mostly at school, they did not hear the parent-to-child speech that is possibly very different from speech addressed to older children or adults. Thus, for the present study, adult-to-adult speech corpora were chosen as a representation of 'classroom English' the children hears in schools. However, the nature of input that L2 children receive certainly is an issue that needs to be addressed in research.

The spoken subcorpus of Mark Davies' Corpus of Contemporary American English (COCA, Davies, 2008) was selected because at present it is the largest corpus of spoken American English available online, and it is the only corpus of American English that is equipped with search tools and is marked up for part-of-speech tags. COCA is a 400-million-word corpus that comprises samples of written and spoken language from 1990 to 2009. The size of the spoken subcorpus of COCA is approximately 81 million words, which makes it the biggest corpus of spoken English (for comparison, the size of the spoken BNC subcorpus is approximately 10 million words). Even though the children in the study had exposure to Canadian English, there is no principled reason to think the relative distribution of definite and indefinite articles and auxiliaries BE and DO would be substantially different in American and Canadian varieties of English.

The frequencies of various inflected forms of auxiliaries BE and DO as well as frequencies of articles were calculated in COCA. In the case of auxiliaries, the frequency of each verb was broken down into frequencies of specific inflected forms, in order to compare the distribution of various person and number forms

(*am*, *are*, and *is*). The frequencies of third person forms were also broken down by subject type (nominal and pronominal), in order to compare the frequency of noun+auxiliary and pronoun+auxiliary collocations. COCA was also used to assess the distribution of definite and indefinite articles in adult English. In addition to raw frequencies of articles in isolation, their frequencies in various syntactic contexts were calculated, such as articles used with nouns in subject and object positions and in existential *there is a N* constructions.

CHAPTER 5. THE ROLE OF L1 IN CHILD L2: THE ARTICLE SYSTEM

5.1 Introduction

In this chapter, elicitation data from children learning English as a L2 are discussed. This chapter focuses on the children's acquisition of the English article system in general and on the role of L1 in particular. The goals of the chapter are the following: to find out whether there is evidence for L1 transfer in the process of child L2 acquisition of English; if there is L1 transfer, to estimate how long the effect lasts; and finally, to compare the impact of other factors on article acquisition that cannot be traced to L1, i.e. developmental factors that may be uniform across all children acquiring L2 English. Data from selected groups of children from two studies were used (see Chapter 4): one is the longitudinal study of a small group of children followed up for two years, and the other one is the cross-sectional study of a larger group of children separated into sub-groups based on their L1. This chapter is organized as follows: the introduction briefly summarizes the findings of previous studies on L2 acquisition of articles that motivated the present study. A detailed discussion of previous work on articles was given in Chapter 3. After the introduction, the relevant properties of the participants' L1s are briefly described (details were given in Chapter 3.3), followed by a discussion of possibilities for transfer of L1 features into L2 English. Section 5.2 presents the method used in the studies and how the article suppliance was coded and analyzed. Sections 5.3 and 5.4 are results sections for the longitudinal and the cross-sectional studies respectively. Finally, Section 5.5 provides conclusions on the role of L1 in children's acquisition of the article system, bringing together the findings of the two studies.

5.1.1 Background and aims of the present study

Numerous studies have been dedicated to the acquisition of article systems in L2 in general and in L2 English in particular. The review of L2 acquisition

literature indicated that while there is agreement that all English language learners have difficulty in using articles (at least initially), there is no consensus as to what the reasons for this difficulty are, and whether these reasons could be the same for all L2 learners. Most of the work on the acquisition of L2 English has focused on adult learners of English with varying degrees of proficiency, and prior research has almost exclusively focused on L2 learners with L1s lacking articles, such as Korean, Russian, Japanese, or Mandarin Chinese (Chapter 3, Section 3.2.3). However, the only way of confirming L1 transfer effects is to compare two or more groups of learners from typologically different language backgrounds. A few recent studies (Hawkins et al., 2006; Snape et al., 2006; Ionin et al., 2008) compared L2 learners from contrastive L1 backgrounds. They found that L1 transfer facilitates L2 acquisition in learners from such language backgrounds as Greek or Spanish (i.e. languages that have the definite/indefinite article distinction) indicated by their native-like performance in L2.

The acquisition of articles has not been systematically investigated in young learners of L2 English, and the majority of L2 studies of children were aimed at investigating their acquisition of English verbal morphology (Chapter 3, Section 3.4.3). There is still no consensus regarding the effect of L1 in child L2 acquisition of various domains of morphosyntax. It was mentioned in the review in Chapter 2 (Section 2.2.1), there are studies reporting evidence for L1 transfer effects at the early stages of child L2 acquisition, as well as studies providing evidence against L1 transfer. There is a need to investigate the issue of L1 transfer in child L2 learners comparing groups from typologically different L1 backgrounds. The analyses in this chapter will focus on the children's accuracy in using articles in obligatory contexts and the types of errors made in such contexts. In addition to L1 influence, other factors will be considered that may affect article use in L2 learners. For instance, it is possible that there is a certain order in which articles are acquired due to their inherent semantic features, and this order is the same across various groups of learners. Thus, there might be commonalities in the performance of groups with typologically different L1s.

5.1.2 Acquisition models and predictions

In the discussion of theoretical approaches to L1 transfer in L2 acquisition in Chapter 2 (Section 2.2.1), it was mentioned that one of the most influential views of this phenomenon in the generative framework is the Full Transfer/Full Access model (Schwartz & Sprouse, 1996). To reiterate, this model hypothesizes that the initial state of L2 acquisition is the entire grammar of the learner's L1 (full transfer) and that in the process of learning the grammar gets restructured, never violating the universal principles (full access to Universal Grammar). The restructuring of the initial L1 grammar happens when it fails to account for the L2 input that the L2 learner receives. Researchers working on acquisition in the generative framework argued that the presence of articles in a language signals the presence of the abstract category of Determiner. If functional projections are initially transferred from L1, DP structure is also expected to be transferred. Following FT/FA, speakers of languages like Spanish or Arabic are expected to transfer the knowledge of the category Determiner from their L1 to English. If a learner's L1 does not have the D category, then that learner will have to establish the presence of this category in L2 from input and add a functional layer to their noun phrases. According to this account, learners from such backgrounds are expected to omit articles at the onset of L2 acquisition, because their initial L2 functional structure (coming from their L1) might not include a D projection, or the necessary semantic features assigned to lexical items like articles. It follows from this model that learners from [+article] L1 backgrounds will be more accurate with articles at the early stages of L2 acquisition.

In the first part of the study, the Full Transfer model will be tested by comparing accuracy and errors in the [+article] and [-article] groups of learners. The general question for the comparison of these two groups is the following: Is there L1 influence in the form of a facilitation effect for the children whose L1s have article systems, and a delayed acquisition in children whose L1s do not have articles? In terms of accuracy with article provision and errors that L2 children can make, specific expectations for L1 effects can be formulated as follows:

(1) Full Transfer/Full Access predictions:

a. Accuracy: [+article] L1 children are expected to be more accurate than the [-article] L1 children

b. Error types: there will be more article omission errors in [-article] L1 children

However, we also need to ask whether there are differences to be expected between groups of learners whose L1 article systems are parallel to the English system (i.e. they have a definite and an indefinite article) and those whose L1s are not completely similar (e.g. Arabic has a definite article but no indefinite article). Such differences cannot be captured by the FT/FA model in its original form since it only analyzes L1 transfer as transfer of abstract categories such as Determiner, rather than specific lexical items. Cross-linguistic differences such as the one between Arabic and English are better captured by a feature-assembly approach (see Chapter 2, Section 2.2.2). The feature-assembly approach provides a formal description of differences among languages in terms of what features they select from the universal inventory and use in the assembly of 'feature bundles' that are mapped onto lexical items. Under a feature-assembly account, L2 learners start off with a fully assembled set of L1 lexical items and grammatical categories. Thus, both the Spanish- and the Arabic-speaking learners are expected to transfer the knowledge of the functional category Determiner to their L2 English, together with the features that are included in it. The difference between the two groups is that the Spanish speakers know that [+definite] and [–definite] features both are mapped onto an overt vocabulary item (the article), whereas the Arabic speakers need to learn that the feature [-definite] is mapped onto the overt indefinite article on singular nouns, rather than a null article. Predictions of this account are summarized in (2):

(2) Feature assembly predictions

a. Accuracy: Spanish and Arabic L1 groups are expected to be more accurate than the Chinese and Hindi L1 groups in definite contexts

b. Error types: Omissions are expected in all contexts for the Chinese group; for the Arabic group, omissions are expected only in indefinite contexts

5.2 Method

Two corpora were chosen as the source of data: a longitudinal study of a relatively small group of children and a cross-sectional study of a large group. For the investigation of the acquisition of articles, both studies were used and subgroups of learners from each corpus were selected. The following section discusses the elicitation materials and coding of elicited data, which were the same in both studies. The rationale for the selection and grouping of participants for the longitudinal and cross-sectional study were explained in detail in Chapter 4 (Sections 4.1.1-4.1.2).

5.2.1 Participants: The longitudinal study

Article systems of the L1s represented in the study were discussed in Chapter 3 (Section 3.3). On the basis of the children's L1 characteristics in this study, in the first part of the study they will be divided into two subgroups, [–article] and [+article] L1s simply based on whether or not the language has an article morpheme of any kind. In the second part, L1 transfer effects will be investigated further by comparing four separate L1 groups in order to find out whether specific similarities or dissimilarities between L1 and L2 affect learners' performance.

The languages of the majority of the children in the [-article] group were Mandarin and Cantonese Chinese, and there was also one speaker of Korean and one speaker of Japanese. Recall that unlike English, these languages do not have an article system. Functions similar to those of the English articles are performed by a combination of devices, such as classifiers, numerals and word order. The languages in the [+article] group were Arabic and Spanish, with one speaker of Romanian also included in the group. These languages have a set of morphemes that developed highly specialized uses to indicate (in)definiteness, and so they were considered as languages with an article system similar to English. In the first

half of the study (Section 5.3), Arabic speakers will be pooled together with Spanish speakers in the [+article] L1 group and children within this group will not be compared. Recall that in the longitudinal study, data were collected every 6 months for two years. Thus, the corpus consists of 5 rounds (See Tables 4.1 and 4.2 in Chapter 4 for the list of participants and more information about their backgrounds).

5.2.2 Participants: The cross-sectional study

In the second half of the study (Section 5.4), the Arabic and the Spanish speakers will be separated into two groups in order to see whether there are more subtle L1 effects due to the differences in the article systems of these languages. An additional group of speakers of Hindi, Urdu and Punjabi will be added, which are all closely related article-less languages (Chapter 3, Section 3.3.4). Thus, there will be four L1 groups in the second part of the study: Arabic, Spanish, Mandarin/Cantonese Chinese and Hindi/Urdu/Punjabi[6]. There will be 10 children in every L1 group, and all children are 5-6 years old, similarly to the longitudinal study. Their average age is 5;9 and their average exposure to English is 10 months (see Chapter 4, Table 4.3 for more details about these four groups).

5.2.3 Materials

For the present study of the children's use of articles, one part of the children's speech corpus will be analyzed, namely the transcriptions of narratives. Picture books used for the elicitation of narratives were designed as a part of the Edmonton Narrative Norms Instrument project (ENNI), which was discussed in Chapter 4 (see Section 4.2.2). The advantage of narratives as a method of speech elicitation is that it allows us to tap the children's ability to present a situation that is unfamiliar to the hearer. An important part of this ability is the appropriate use of articles. Since all the narratives were based on the same picture sets, it was possible to control children's use of referring expressions and compare narratives across children. Narratives are useful in testing children's knowledge of articles

[6] In tables and figures these groups will be referred to using general labels 'Chinese' and 'Hindi'

because it is possible to clearly distinguish utterances where the new characters are introduced and the utterances where they are part of the common knowledge, which is precisely the contrast encoded in the indefinite-definite article contrast in English.

5.2.4 Coding procedure

The analysis was limited only to referring expressions that were used for the characters and concrete objects that were crucial for the storyline, i.e. four animate characters and three objects in each of two picture books used for elicitation. Plural, mass and uncountable nouns, as well as idiomatic use of articles in expressions such as *to have a headache* or *to go home*, were excluded. The analysis concerned only the use of articles and did not mention other ways to refer to characters, such as personal and deictic pronouns and proper names (e.g. *someone*, *Missis Doctor*, *this woman*, *his mother*). In sum, the analysis was limited to the use of articles with singular countable nouns because these nouns required an article in all contexts.

The coding rationale was described in detail in Chapter 4 (Section 4.2.2). The nouns that met the inclusion criteria described in this section, there were two kinds of obligatory contexts, indefinite and definite, and three possible codes for article use in each context:

Indefinite: - correct (*a*+noun, e.g. *He had a plane*)
- commission (*the*+noun, e.g. *He had the plane*)
- omission (bare noun, e.g. *He had plane*)

Definite: - correct (*the*+noun, e.g. *The plane fell into the pool*)
- commission (*a*+noun, e.g. *A plane fell into the pool*)
- omission (bare noun, e.g. *Plane fell into the pool*)

Since it is relatively easy to distinguish indefinite and definite contexts in storytelling, this coding procedure was straightforward and objective, as opposed to coding spontaneous speech where the coder would have to make a subjective

judgment on the semantic status of most nouns in each utterance. For the analyses of the children's use of articles, only data from the storytelling talk were be included.

5.3. Results I: The longitudinal study

This section presents the results of the analysis of the narratives elicited from the children in the two-year longitudinal study. The children are separated into two groups according to the typology of their L1, and their article provision and omission in obligatory contexts is analyzed. The [+article] and [-article] groups are contrasted with respect to their accuracy with article use (Section 5.3.1) and with respect to error types (Section 5.3.2). Statistical analyses of the children's performance are performed in order to ascertain the effect of the L1 typology and the amount of exposure to English, as the children's performance improves in the course of the two-year study. The results are summarized in Section 5.3.4.

5.3.1 Accuracy in obligatory definite and indefinite contexts

The question asked for the analyses in this section was: Are the rates of acquisition different for children in the [–article] than in the [+article] L1 group? Children's accuracy in indefinite contexts was calculated as a percentage of all obligatory contexts in which a was supplied, obligatory contexts being singular nouns mentioned for the first time that were not proper nouns. Children's accuracy in definite contexts was calculated as a percentage of all obligatory contexts for *the* in which *the* was supplied. Individual accuracy scores used to calculate the means are given in Appendix 1. The resulting mean percent correct use in context of *a* and *the* for each round are shown in Figures 5.1a and 5.1b, for the [+article] and [–article] L1 groups separately (bars in the graphs represent standard errors).

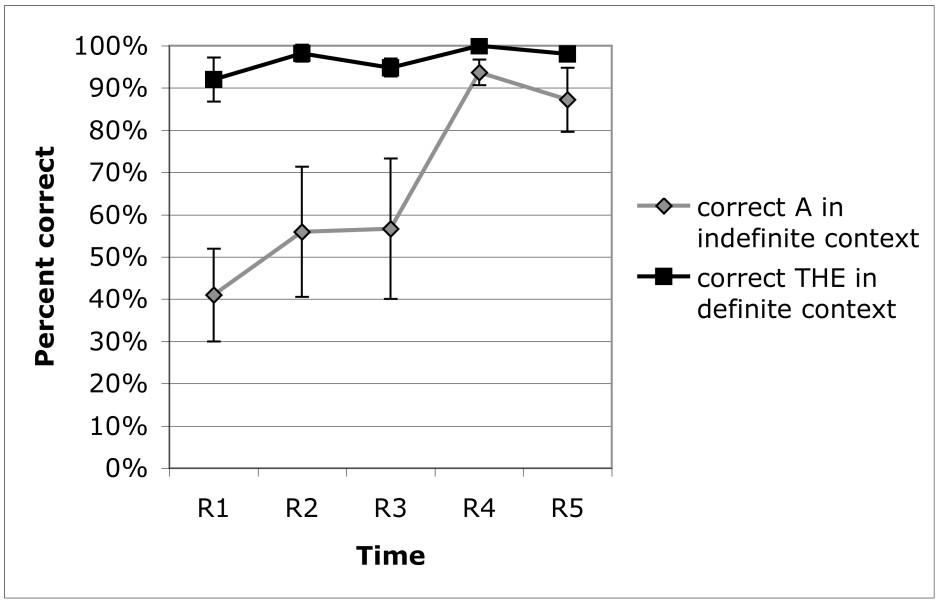

Figure 5.1a. Mean percent correct use of articles for [+article] L1 Group

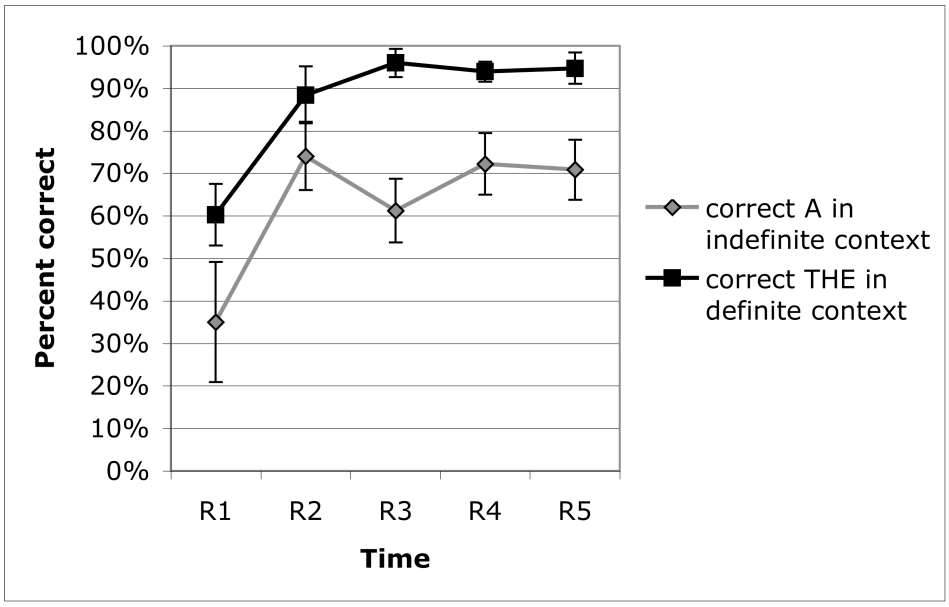

Figure 5.1b. Mean percent correct use of articles for [–article] L1 Group

Mixed ANOVA with L1 background as a between-subjects factor ([+article] group and [–article] group) and round (5 levels) and article type (definite and indefinite) as within-subjects factors. Both round ($F(4,48) = 10.68$, $p = .000$, partial eta squared = .471) and article type ($F(1,12) = 31.723$, $p = .000$, partial eta

squared = .726) main effects were significant, but L1 background was not (F(1,12) = 2.030, p = .180, partial eta squared = .145). Significant linear trends for both round (F(1,12) = 30.882, p = .000, partial eta squared = .720) and article type (F(1,12) = 31.723, p = .000, partial eta squared = .726) confirm that the learners grew steadily more accurate over time with their use of articles. The significant main effect for the two-level factor, article type, confirms that the children were more accurate with the definite than indefinite article, as the means for the former were consistently higher.

As for the interactions of the factors, the ANOVA yielded one significant interaction: round x article type x L1 background (F(4, 48) = 2.914, p = .031, partial eta squared = .195). To further explore this interaction, I conducted post-hoc independent sample pairwise t-tests on the correct use of each article type between each L1 group at each round. Applying a Bonferroni correction to the alpha level to control for Type I error (alpha = .005), we found just one significant result in all the pairwise comparisons: The [–article] group had lower accuracy with the definite article in context at round 1 than the [+article group] (60.2% vs. 92.2%, t(13) = -3.869, p = .002).

It can be concluded from this post-hoc analysis that over all, L1 background did not exert much influence on children's acquisition of articles with the exception that the [–article] group were lagging behind in accuracy at round 1 with *the*. Put differently, article type was a more important factor than L1 background in acquisition patterns and rates. Children in both groups were more accurate with *the* in definite contexts than with *a* in indefinite contexts throughout the two year period of observation, and acquired the use of *the* in definite contexts by round 3.

5.3.2 Error types and the role of L1

The questions asked for the analyses in this section were: (1) what is the relative distribution of commission and omission among children's errors?; and (2) does L1 background interact with error types, for example, do the [–article] children have more article omissions than the [+article] children?

The frequencies of error types per child per round varied such that in some cases they were less than 2. Therefore, analyses in this section were designed such that percentages per child, per round were not calculated on the grounds that they could be unreliable. First, the proportion of incorrect use of *the*, *a* and null articles was calculated from the total number of contexts at each round across all children, and divided by L1 group and by definite and indefinite context. For instance, percent incorrect use of *the* in indefinite context was the number uses of *the* in indefinite contexts divided by the total number of indefinite contexts. There were four possible types of errors: *the* in indefinite context, *a* in definite context, and null article in indefinite context, and null article in definite context. (Recall that null articles in this story-telling task would always be errors, even though in many semantic contexts in English, null articles are a grammatical choice). The results of these analyses are presented in Figures 5.2a and 5.2b (the numbers used to calculate the percentages are given in Appendix 2).

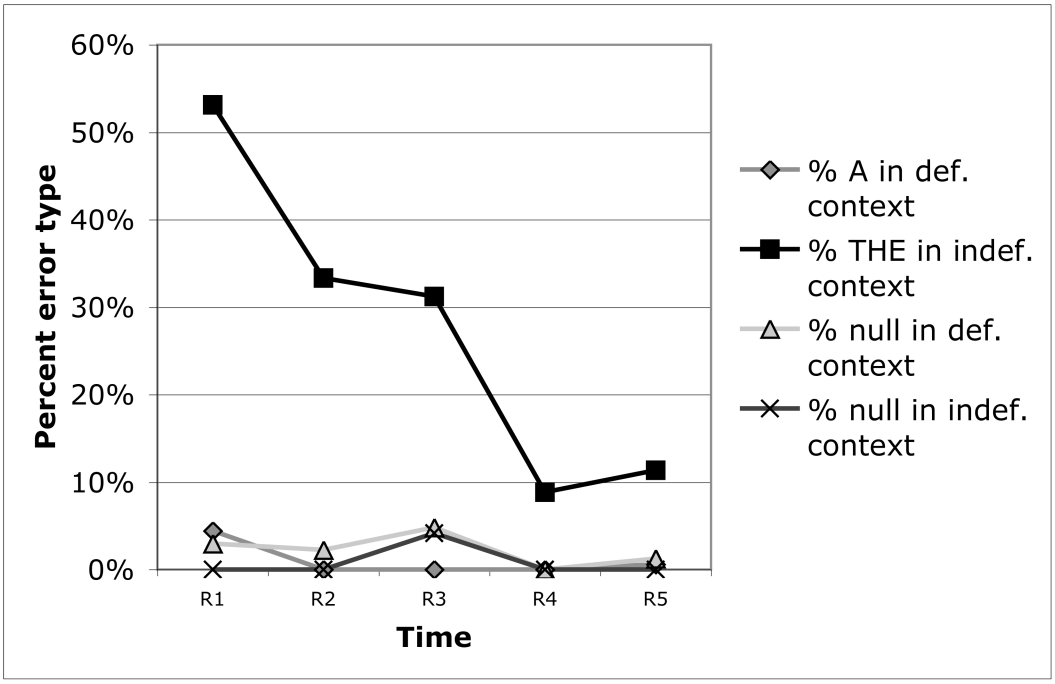

Figure 5.2a. Percent distribution of error types for [+article] L1 Group

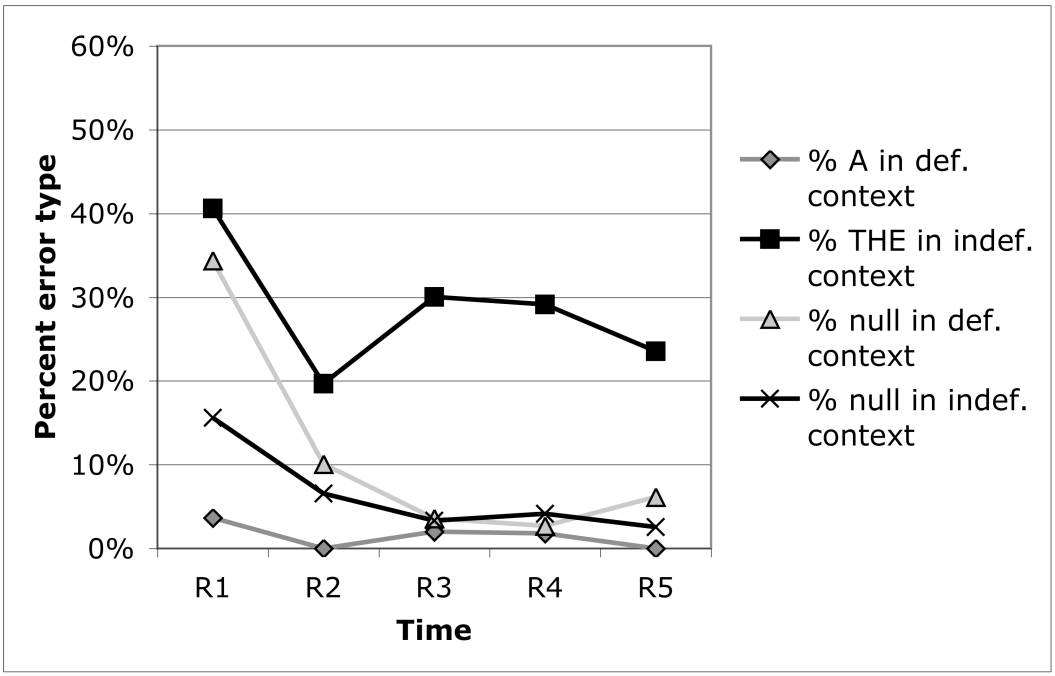

Figure 5.2b. Percent distribution of error types for [–article] L1 Group

The results demonstrate that *the* misuse was clearly the dominant error type for both the [+article] and [–article] groups, and that null articles were an error type specific to the [–article] group, since they were negligible in the [+article] group data but appeared in the [–article] group data in both definite and indefinite contexts. Furthermore, null articles began to disappear even for the [–article] group after round 2, indicating that these errors were more frequent during the early stages of these children's English L2 acquisition.

To complement this distributional analysis, independent sample *t*-tests were conducted on children's mean proportions of the different error types across all rounds, divided by context and L1 group. The mean proportions of incorrect *a* and null articles in definite contexts, and incorrect *the* and null articles in indefinite contexts, for the [+ article] and [–article] L1 groups are given in Table 5.4, with the results of the *t*-tests. The *t*-test analysis supports the distributional data in Figures 5.2a and 5.2b: the only significant group differences are for null article use in definite and indefinite contexts, with the [–article] children having a higher proportion of these errors.

Table 5.1

Mean proportions of error types by group

	L1 Group	Mean	*t*-value	*p*-value
Incorrect *a* in definite context	[–article]	.013		
	[+article]	.010	$t(15) = 0.477$	$p = .641$
Null article in definite context	[–article]	.088		
	[+article]	.029	$t(15) = 2.375$	$p = .036$*
Incorrect *the* in indefinite context	[–article]	.281		
	[+article]	.249	$t(15) = 0.459$	$p = .653$
Null article in indefinite context	[–article]	.035		
	[+article]	.009	$t(15) = 2.638$	$p = .019$*

5.3.3 The rate of acquisition and convergence

The children in this longitudinal study were expected to show improvement in their accuracy with articles, with the [–article] children being slower due to the absence of facilitative transfer of the category D from their L1s. Accuracy levels in context of 90% or higher were achieved early on for *the* by both groups, and for *a* at round 4 by the [+article] group. Comparing the results from child L2 learners in our study and adult learners of L2 English in previous studies, it can be concluded that the child learners certainly converged faster. For instance, Hawkins et al. (2006: 17) reported 50%-58% *the* misuse in the advanced Japanese group, whereas for the children in our study, at round 5 (34 months of exposure) there was about 10% *the* misuse in the [+article] group and about 20% *the* misuse in the [–article] group, and in any case, no significant between-group differences emerged at round 5. Regarding the predicted slower development of the [–article] group, the lower accuracy with *the* in definite context at round 1, and the absence of ceiling scores for use of *a* in indefinite context by the end of the study, suggest that the [–article] group displayed slower acquisition rates, but only to a limited extent.

5.3.4 Summary: The longitudinal study

Are there L1 effects? With respect to accuracy with articles, statistical tests revealed that the only significant difference between the two groups was in round 1 when the [–article] group were lagging behind in accuracy with *the*. Overall, L1 background did not exert much influence on children's acquisition of articles, as article type was a more important factor than L1 background in acquisition patterns and rates.

Does L1 background interact with error types? Under Full Transfer, the [–article] L1 group, but not the [+article] L1 group, was expected to omit articles. As predicted, the [–article] L1 group had null article errors early on, in both definite and indefinite contexts, while the rates of omissions were negligible in the [+article] L1 group. A possible interpretation of this difference is that the absence of the category D[7] in the initial state grammars of the [–article] group was reflected in their omission errors in language production. Article omission rates began to drop for the [–article] group after round 2, indicating that these errors were a characteristic of the early stages of these children's English L2 acquisition.

Developmental patterns: Accuracy Both the [+article] and [–article] L1 groups were expected to be more accurate with *the* in definite contexts than with *a* in indefinite contexts. As predicted, I found that the target article was used more often in definite contexts than in indefinite contexts. This difference was present at all stages of acquisition, and across learners with different L1 backgrounds, and thus, was a robust pattern.

Developmental patterns: Overuse of THE While null articles were an error type specific to the [–article] group, *the* misuse was clearly the dominant error type for both the [+article] and [–article] groups. Thus, [+article] L1 learners did not seem to transfer the knowledge of the indefinite article from their L1. Under a UG account, I can assume that learners in both groups had access to the inventory of semantic features, such as [+/-definite] and [+/-specific], as well the inventory of functional projections including D, from Universal Grammar. Perhaps at the

[7] The generative model of L2 acquisition is unclear as to when a category is considered to be acquired. Therefore, the criterion of 'having' a category was set to be 90% accuracy, which would indicate that the learner's use of the relevant morpheme is stable.

outset of acquisition, the L2 learners from [+article] L1 backgrounds transferred the knowledge of the category D from their L1 functional structure, but did not necessarily transfer the grouping of semantic features from their lexicons. Such an explanation diverges from the original formulation of the FT/FA, which states that at the initial stage of L2 acquisition all the syntactic properties of the L1 are transferred into the L2 grammar, excluding the surface phonological realizations of morphemes. In the next section, the early stage of L2 acquisition in children will be discussed in terms of feature assembly, which might provide a better description for the limited transfer.

The effect of the length of exposure to English Since young L2 learners were not expected to be influenced by age-of-onset effects, it was not surprising that the learners in this study improved in the course of the two-year data collection, with the initial delay in the [–article] children probably being due to the absence of transfer of the category D from their L1s. However, evidence for significant improvement was not found in all contexts, since accuracy levels of 90% or higher were achieved early on for *the* by both groups, and for *a* at round 4 only by the [+article] group. Child L2 learners converged on the target system faster than what prior reports indicated for adult L2 learners. For instance, Hawkins et al. (2006), Lardiere (2004) and White (2003) studied adults from [–article] backgrounds who had been learning English for over five years, and found that they had not yet converged on the target article system. In sum, the children's underlying grammatical competence was influenced by L1 at the earliest stages of L2 acquisition (in line with FT/FA), since L1 transfer is likely to be the reason for the lower accuracy rates at Round 1 for the [-article] group, and the significant difference between the article omission rates in [+article] and [–article] groups.

5.4 Results II: The cross-sectional study

The first part of this chapter tested predictions of the Full Transfer/Full Access model. In the second part of the chapter, L1 transfer patterns are explored further in terms of feature assembly, using new groups of participants, which are larger and more homogeneous with respect to L1 backgrounds. The expectations for L1

transfer in terms of feature assembly model were as follows: Spanish and Arabic L1 groups were expected to be more accurate than the Chinese and Hindi L1 groups in definite contexts. Omissions were expected in all contexts for the Chinese group. For the Arabic group, omissions were expected only in indefinite contexts

This section presents the results of the analysis of the narratives elicited from the children in the cross-sectional study. The methodology and the rationale for data coding were exactly the same as in the longitudinal study. For this follow-up analysis the children were split into four language-specific groups with 10 participants in each group (Arabic, Chinese, Hindi and Spanish). The analyses in this section follow the same logic as those in the longitudinal study: First, children's accuracy with article suppliance is analyzed (Section 5.4.1), followed by error types (Section 5.4.2). Statistical analyses of the children's performance are performed in order to ascertain the effect of the L1 typology and the amount of exposure to English. The results are summarized in Section 5.4.4, followed by an overall summary of the findings in this Chapter (Section 5.5)

5.4.1 Accuracy in obligatory definite and indefinite contexts

In order to test the first prediction in (22), the accuracy of the provision of the definite and indefinite article will be compared across the four L1 groups. Accuracy with *the* was calculated as a percentage of all obligatory definite contexts in which the target article the was supplied. Accuracy with *a* was calculated as a percentage of all obligatory indefinite contexts in which the target article *a* was supplied. A percent correct score was calculated for every child (see Appendix 3 for individual scores), and the resulting mean percent correct scores are given in Figure 5.3.

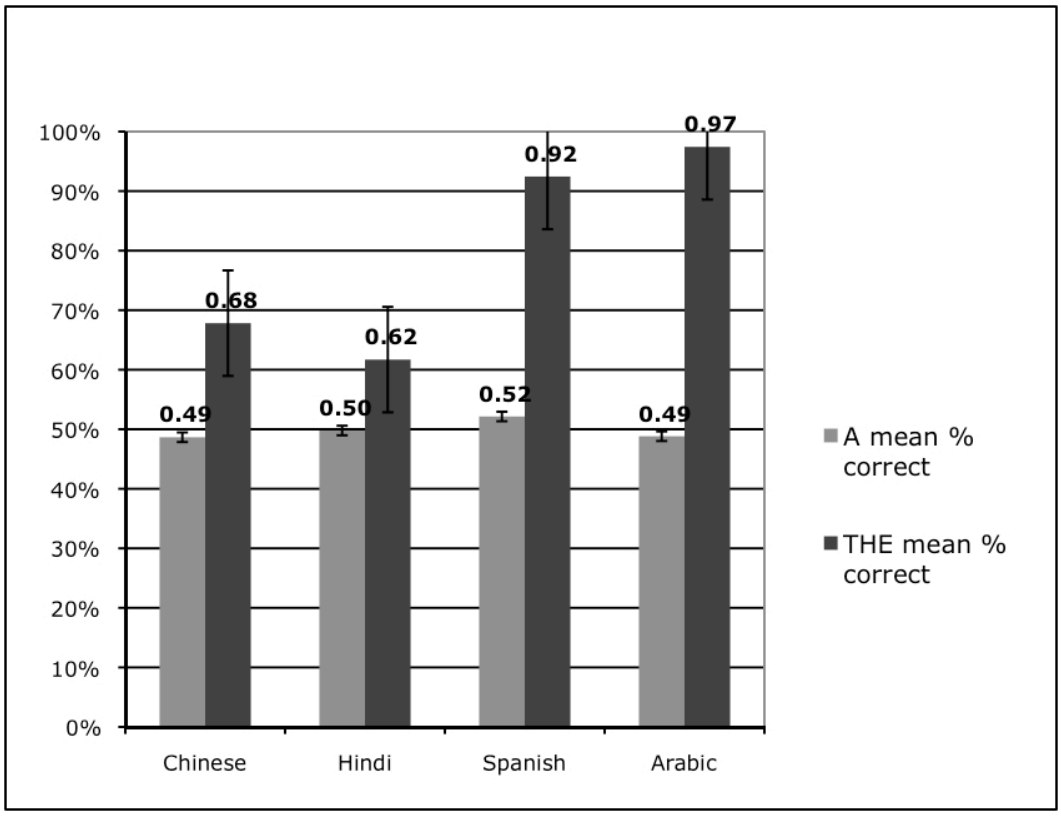

Figure 5.3. Mean percent correct for production of articles in obligatory contexts

It is apparent from the distributional analysis in Figure 5.3 that the accuracy with the indefinite article was very similar in the four groups. The Spanish-speakers did not outperform the three other L1 groups despite the presence of an indefinite article in their L1. The L1-based effect emerges in the accuracy rates in definite contexts. They are very similar in the Chinese and the Hindi groups on the one hand, and the Spanish and Arabic groups on the other, with the latter pair being more accurate than the former pair. In order to confirm the L1 effect in the children's accuracy, a one-way ANOVA was conducted with L1 as a between-participants factor (4 levels) and accuracy with definite articles as the dependent variable. The ANOVA was significant ($F(3,36) = 4.887$, $p = .006$). Follow-up independent samples *t*-tests were carried out to confirm differences between the means. Applying a Bonferroni correction to the alpha level to control for Type I error ($\alpha = .0083$), it was revealed that the difference between mean accuracy scores reached significance only in the Hindi and Arabic L1 groups comparison,

with the Hindi speakers having the lowest proportion of target article use (.62), and the Arabic speakers having the highest proportion (.97) in definite contexts.

Table 5.2

Pairwise comparisons of mean accuracy scores in definite contexts

L1 pairs		Mean difference	*t* value	*p* value
Chinese	Hindi	.062	.398	.695
	Arabic	-.295	2.937	.009
	Spanish	-.245	3.329	.032
Hindi	Arabic	-.357*	2.936	.008
	Spanish	-.307	2.466	.024
Spanish	Arabic	-.050	1.361	.19

Thus, some L1 effects emerged in definite contexts, but since all groups performed uniformly in indefinite contexts, it can be concluded that overall, the effect of L1 was a weaker factor than article type, echoing the results of accuracy analyses in the first half of the chapter. Recall that accuracy rates with *the* were higher than with *a* in all children in the longitudinal study throughout the two-year period of observation. Thus, the acquisition of the definite article before the indefinite article appears to be a robust pattern in child L2 acquisition.

Accuracy and months of exposure

One of the findings of the longitudinal study reported in the first half of the chapter was that L1-based differences in accuracy were observed only in round 1 of data collection (average 9 months of exposure). Since the amount of exposure to English in the groups selected for the present analysis ranged from 2 to 18 months, it was necessary to determine whether it was also the case that the children who had had more months of exposure to English were consistently more accurate with articles. To complement the cross-group comparisons of accuracy reported in the previous section, Pearson correlations between children's MOE and accuracy scores were calculated, first pooling together children from all four

L1 groups. Due to the size of the group (40 participants), it was possible to use a parametric test. It was decided not to split the participants into groups based on MOE, as there was no independent criterion for a cut-off point. Thus, children's amount of exposure was included in this statistical analysis as a continuous variable. The correlation between the amount of exposure and percent correct score with *the* was significant at the .05 level ($r = 0.364$, $p = 0.021$). There was no significant correlation between amount of exposure and percent correct scores with *a* ($r = 0.055$, $p = 0.738$). In other words, for the children in this study, the suppliance of *the* in obligatory contexts significantly improved along with their length of exposure, but the suppliance of *a* in obligatory contexts did not.

5.4.2 Error types and the role of L1

This section further explored the question of whether L2 children are influenced by the knowledge of L1 in the acquisition of the English article system. Similarly to the previous study, article omissions were expected in all contexts for the Chinese group and omissions in indefinite contexts for the Arabic group. In order to check for L1 influence effects in the data, a mean score for each error type for each L1 group was obtained. The error scores were calculated as follows. In obligatory indefinite contexts, the proportion of definite articles (commission error) and the number of bare nouns (omission error) was calculated. In obligatory definite contexts, the proportion of indefinite articles (commission error) and the number of bare nouns (omission error) was calculated (recall that we excluded plural nouns and names from the analysis, thus bare nouns were always omission errors). Thus, four error scores were calculated for each child: percent overuse of *the* in indefinite contexts, percent omission of *a* in indefinite contexts, percent overuse of *a* in definite contexts, and percent omission of *the* in definite contexts (see Appendix 3 for individual scores). The mean error scores for each L1 group are presented in Figure 5.4.

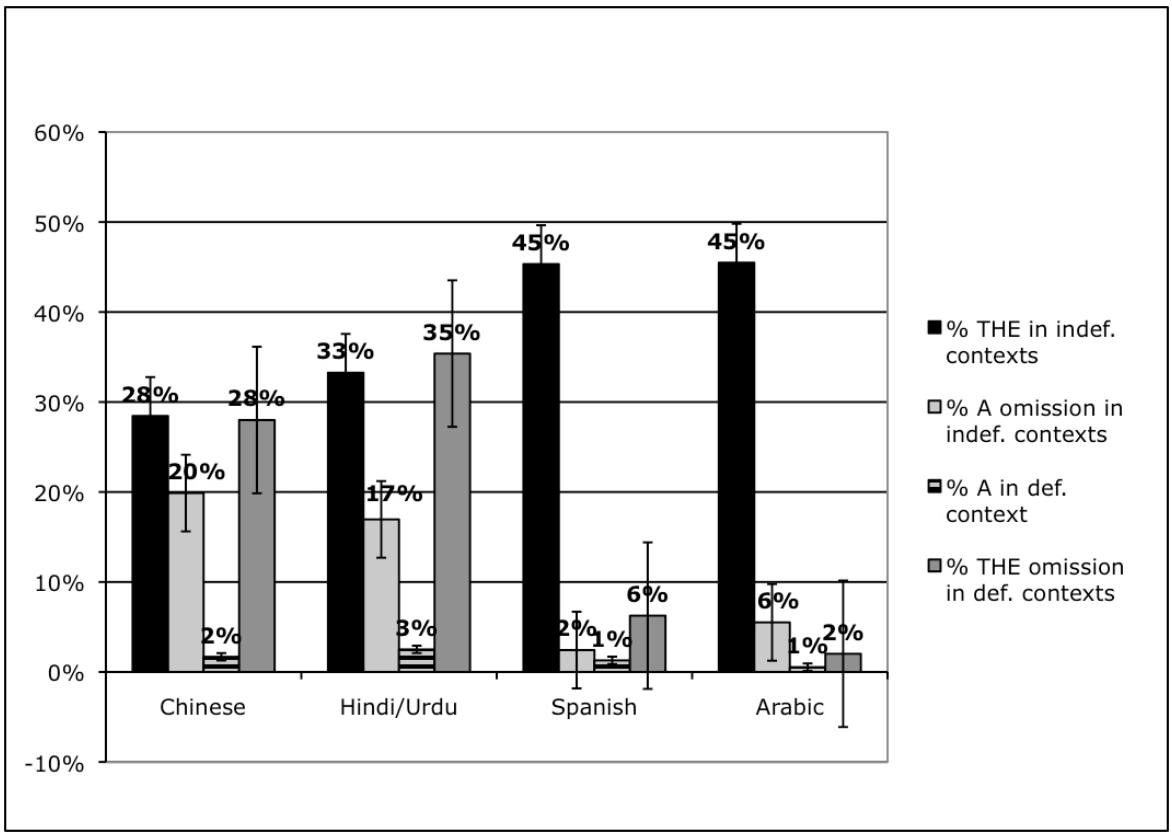

Figure 5.4. *Mean percent commission and omission errors in indefinite and definite contexts*

This distributional analysis demonstrates that the rate of *the* substitution in indefinite contexts was high in each L1 group. The profiles of error distributions are clearly similar in the first two groups (Chinese and Hindi), and in the third and the forth groups (Spanish and Arabic). The difference in the patterns across the groups is due to the relatively high proportion of omission errors in the Chinese and Hindi groups, both in indefinite and definite contexts, which was predicted. Also as predicted, there were very few omissions in the Spanish and Arabic L1 groups, and the two groups showed a very similar error pattern in general. In fact, in these two groups, the rate of *the* substitution was the same (*the* was erroneously supplied in 45% of all indefinite contexts). The rate of omissions in these two groups was very low (up to 6%), especially compared with omissions in the Chinese and Hindi groups (up to 35%).

In the analysis of error types by L1 background, it was not possible to apply a parametric statistical test due to a large number of categories between subjects and small numbers of subjects in each group (10 for each L1 background). Thus, non-parametric Wilcoxon Signed Rank Tests were performed on the main trends in the error distribution, comparing selected pairs of proportions for omission and overuse of *the*. These tests supported the observations based on the descriptive data in Figure 5.4. In the Chinese and Hindi groups, there was no significant difference between the rate of *the* omission in definite contexts and *the* overuse in indefinite contexts (z(Chinese) = -0.306, p = 0.760; z(Hindi) = -0.153; p = 0.878). In the Spanish and Arabic L1 groups, the rate of *the* omission was significantly lower than the rate of *the* overuse (z(Spanish) = -2.599, p = 0.009; z(Arabic) = -2.803, p = 0.005). Thus, the Spanish and Arabic L1 children were more likely to substitute *the* in indefinite contexts than commit any other type of error.

It was reported above that there was a significant correlation between the amount of exposure and percent correct score with *the* but not with *a*, partially supporting our prediction. To follow up on the findings regarding error patterns that brought out significant rates of omission in the Chinese L1 group and the Hindi L1 group, it was decided to further investigate L1 influence in these two groups and see whether the correlations between the amount of exposure to English and children's accuracy were significant in each L1 group. Non-parametric Spearman correlations were calculated between MOE and percent correct in indefinite and definite contexts. Accuracy with *the* was significantly correlated with MOE only in the Chinese L1 group (r = 0.758, p = 0.011) and the Urdu L1 group (r = 0.652, p = 0.041). Accuracy with *a* was not significantly correlated with MOE in any of L1 groups. This finding is not surprising if we recall from the previous analysis that the Chinese and Hindi L1 groups frequently omitted *the* in obligatory contexts, thus having room for improvement, while the Spanish and Arabic L1 groups made very few errors in definite contexts.

5.4.3 Accuracy and the role of syntactic context

A pattern emerged in the process of data analysis that did not follow from the predictions of the models discussed in this chapter. In both the longitudinal and the cross-sectional studies, the children appeared to supply the indefinite article correctly in some syntactic positions more frequently than in others. The children often used an indefinite in article accurately when the target noun was introduced as an object, e.g. *the dog was building **a** castle*. In the subject position, nouns were often introduced erroneously with a definite article. Furthermore, in 'there is' constructions (e.g. *once upon a time there was **a** rabbit*), the indefinite article was never omitted or substituted with *the*, and this pattern was present in children from all L1 backgrounds. Further investigation of syntactic contexts of article use in L1 acquisition of English is needed in order to establish whether the syntactic position of the noun made a difference in the choice of articles. This issue will be addressed in Chapter 8, which will discuss the role of syntactic constructions and input frequency in the distribution of definite and indefinite articles in L2 children in terms of the usage-based framework, which is better suited for the analysis of these issues.

5.4.4 Summary: The cross-sectional study

Are there L1 effects? With respect to accuracy with articles, the only L1-based difference apparent in the data was the difference in accurate suppliance of definite articles, with the Chinese and Hindi groups being less accurate than the Spanish and Arabic groups. No accuracy differences were found between the Spanish-speaking and the Arabic-speaking groups in definite contexts (which was expected), but also there were no differences in indefinite contexts. The latter finding was not expected based on the properties of the article systems of these languages, but it provides support for the findings of the longitudinal study that also did not find L1 transfer of the indefinite article in the [+article] group.

Does L1 background interact with error types? Evidence for L1 influence was found in the Chinese and Hindi L1 groups in the form of article omission errors. At the same time, all children had difficulty with choosing the correct

article form in indefinite contexts, even the Spanish-speaking children. It appears that the Spanish L1 children were supported by L1 transfer only in learning the DP structure, but not in mastering the semantic distinctions between indefinite and definite contexts. In other words, these learners appeared to be aware that singular countable nouns in English must be preceded by an article, since they hardly used any bare nouns. But even though they used an article in an appropriate position, they overgeneralized the use of the definite article to some indefinite contexts. Since this use of the definite article is inappropriate both in Spanish and in English, this error cannot be explained by L1 transfer. Limited L1 support in the structural domain of DP can be further confirmed by the patterns in L1 Arabic children. The Arabic L1 group did not perform worse than L1 Spanish children, despite the nature of their L1, which only has a definite article and no indefinite article. As the rate of indefinite article omissions was negligible in the Arabic and Spanish groups (with the totals of 6/115 and 4/113 omitted articles respectively), it is very unlikely that any differences between them were due to L1 influence. For instance, omissions of *a* in the Arabic group cannot be attributed to the fact that Arabic has no indefinite article, because there was approximately the same (very low) rate of omissions in the Spanish group, and Spanish has both an indefinite and a definite article.

In sum, the four L1 groups in the study formed two contrasting pairs, with both the accuracy patterns and the error patterns being almost the same in the Chinese and Hindi groups on the one hand, and the Spanish and Arabic groups on the other. This finding is interesting since based on L1 differences, a three-way opposition was expected for the Chinese and Hindi vs. Arabic vs. Spanish groups.

Developmental patterns: Accuracy. All L1 groups were expected to be more accurate with *the* in definite contexts than with *a* in indefinite contexts. As predicted, it was found that the target article was used more often in definite contexts than in indefinite contexts. This finding again indicates that what is transferred appears to be merely the knowledge of the functional projection D, but not the particular mappings of feature clusters onto morphological forms. It appears that specific transfer of the mapping of the feature [-definite] did not take

place in the Arabic L1 group. This is perhaps not surprising if we compare this finding with the finding that the Spanish L1 group did not transfer their knowledge of the indefinite article into English either. I have already suggested that the indefinite feature mapping is inherently harder to acquire in both L1 and L2 acquisition (see Summary of the previous study). If the Spanish and Arabic children did not have a stabilized knowledge of this mapping in their L1, they had no pre-assembled feature-morpheme mapping to transfer to their L2. Hence, at the onset of acquisition they were similar to L1 children in this aspect of morphosyntax.

Developmental patterns: Overuse of the. While null articles were an error type specific to the [–article] group, *the* misuse was clearly the dominant error type for all L1 groups. Together with the established higher accuracy with *the* than with *a*, the predominance of *the* substitution errors constituted a developmental pattern, i.e. similar to the patterns reported in previous studies of L1 acquisition of English. Thus, generalizing over the research done on the acquisition of the English article system (discussed in Chapter 4), it can be said that the pattern in child L1, child L2 and adult L2 appears to be uniform, namely that all learners tend to overextend the use of the definite article to indefinite contexts. It was mentioned in the discussion of the article system in terms of semantic features (Chapter 3, Section 3.6.2) that one possible reason for *the* overuse could be that the definite article is less complex than the indefinite article, at least with respect to its featural makeup and contexts of insertion. If featural complexity is the underlying reason for the acquisition order of *the* and *a*, then L1 children are expected to display this pattern and L2 learners are expected to display this pattern as well, regardless of L1 background, which was the case in the child L2 learners in the study.

L1 Influence and the amount of exposure to English. The increase in the months of exposure was associated with improvement in the accuracy with *the*, but there was no significant association between months of exposure and the accuracy with *a*. In other words, the indefinite article use proved so problematic for children that in the first 1.5 years of acquisition there was no significant

improvement in their accuracy in indefinite contexts. It ties in with the finding in the longitudinal study that children had more difficulty using *a* than *the* at all stages of the 2-year study. Generally speaking, in indefinite contexts, accuracy scores were low for two reasons, the overuse and omission. In definite contexts, the only considerable source of error was omission. Thus, children's accuracy in definite contexts significantly improved because article omission was not a protracted pattern.

Furthermore, the improvement in *the* suppliance was significant only in the Chinese and Hindi L1 groups. This finding is not surprising if we take into account that there were many more article omission errors in these two groups than in the Spanish and Arabic L1 groups, as discussed above. This finding can be interpreted as a reflection of L1 influence in these children: the Spanish and Arabic speakers were already over 90% accurate in definite contexts and simply did not have much room for improvement, whereas the other two groups were initially 'slowed down' by having to acquire the functional category D. This finding is consistent with both the feature-assembly account according to which the learning process is expected to be more difficult for learners whose L1s do not assemble the relevant features into functional categories or assemble them in a way that is different from L2. However, bearing in mind that I found significant improvement over 18 months of exposure, our study further demonstrated just how short-lived L1 effects are in child L2 acquisition, in line with the previous studies that also found transient L1 effects in child L2 acquisition.

5.5 Conclusions: The role of L1 transfer in the two studies

What is transferred? Two versions of a generative approach to L2 acquisition were tested: the Full Transfer/Full Access model and the feature assembly model. Predictions of the Full Transfer model regarding the transfer of abstract functional structure of the noun phrase were confirmed in both studies. An interesting generalization emerging from the findings of the two studies taken together is that the L1 typology can indeed facilitate the acquisition of the article

system, but only for the general awareness of the obligatoriness of articles. The particular *type* of the article system in the background of L2 learners did not seem to have an effect, at least in the case of Spanish- and Arabic-speaking children in this study. This finding confirms, in retrospect, that pooling together speakers of these languages did not conceal potential differences within the [+article] group in the longitudinal study. Particular differences between these L1s and English did not seem to matter for the development of the article system in child L2 learners of English. Using the terminology of the feature-assembly approach, the children did not transfer particular feature-morpheme mappings from L1, but rather transferred a general knowledge of an abstract functional layer in the NP structure.

Why is the *'easier' to learn and why is it substituted for* a*?* Differential difficulty by article type does not follow directly from the Full Transfer predictions. However, the feature-assembly approach can be applied to account for the difficulty that learners in both studies had with the mastery of the indefinite article, regardless of their L1 background. In a discussion of article acquisition in an adult L2 learner, Lardiere (2004: 335) suggested that differences in accuracy in indefinite and definite contexts could be due to the fact that "definite articles in English need not take number and the count/mass distinction into account, which makes them less featurally complex than indefinites in at least one respect". A similar conclusion can be drawn from Hawkins et al.'s (2006: 20) discussion of featural contexts of insertion for articles: *a* is inserted in [D, -definite, +singular] contexts, *the* is inserted in [D, +definite] contexts. Thus, in order to use *a* appropriately, learners have to identify the feature [singular] as relevant for the insertion of this article.

Alternative explanations for the substitution of the: One of the factors that need to be mentioned here is the role of phonological properties of articles that could have interfered in the analysis. The acoustic difference between the definite and the indefinite article when they are unstressed is small, and thus could have resulted in transcription errors. Transcription could also be influenced by the word that preceded the article. For instance, if the word that precedes the article ends in

an interdental stop (e.g. *he found **th**e ball*), it could be difficult to tell if the child pronounced the initial interdental fricative in *the*. It is also possible that the child deleted the interdental fricative in the *d-th* sequence leaving just the vowel. Thus, acoustics or articulation can potentially influence the transcription of articles that immediately followed words ending in consonants. However, these factors were unlikely to have a substantial influence on the results. First, obviously very few NPs followed words that ended in interdental stops, especially because children produced very few past-tense *–ed* endings on verbs at the early stages of acquisition. Secondly, the difficulty with transcription was addressed in the methodology (Chapter 4, Section 4.2): a reliability check was done on storytelling and spontaneous speech transcripts, and the inter-transcriber reliability score was 95%.

The overuse of the definite article in indefinite contexts can also be the result of external factors. For instance, the children could see the pictures and they could have assumed that the experimenter also knew who the characters were. In L1 acquisition, *the* overuse has been explained in terms of the 'egocentricity' of very young children, i.e. their failure to take the listener's perspective into account (Schaeffer & Matthewson's, 2005). However, Schaeffer & Matthewson found that 'egocentric errors' disappeared by the age of 4;0. It is unlikely that *the* overuse was due to egocentricity given the age of the children in the study. Another counterargument to this type of explanation is that the children clearly improved over time. If the overuse of articles had been a task effect, it would have stayed more or less constant throughout the longitudinal study.

A slightly different account that is perhaps more applicable to L2 children was suggested for *the* substitution errors in monolingual children by Krämer (2005) and De Cat (in press). Krämer and De Cat suggested that over the age of 4, the underlying knowledge of information structure is in place, but children fail to mark some referents as new to the hearer because they continue to rely on deixis, especially in the presence of visual stimuli. This approach considers *the* substitution to be not a cognitive error, but a discourse error. Krämer (2005) and also found that integrating sentences into coherent discourse was a difficult task

for monolingual children below age 5 or 6. Building a coherent discourse is possibly difficult even for children above the age of 5 or 6 – when they have to carry out this task in a L2. However, while the presence of visual stimuli and difficulty with discourse-building can contribute to *the* overuse, they are insufficient to account for context-specific errors. Most errors of *the* substitution occurred when the target indefinite noun was the subject of the sentence, while existential constructions and object positions were the contexts where the indefinite article was often used by the children correctly, suggesting the influence of the syntactic context on the children's choice of articles. This does not necessarily imply, however, that they used *a* without understanding its semantic features, because *a* did occur in subject positions from Round 1 on, and *a* was not overused in definite contexts. However, differences in accuracy in various syntactic contexts deserve a more detailed investigation, which will be pursued in Chapter 7.

CHAPTER 6. THE ROLE OF FIRST LANGUAGE IN CHILD L2: THE AUXILIARY SYSTEM

6.1 Introduction

This chapter discusses elicitation data from children learning English as a L2. It investigates the children's acquisition of the English auxiliary system, focusing on the auxiliaries BE and DO. The goals of the chapter are similar to those in the previous chapter: to find out whether there is evidence for L1 transfer in the process of child L2 acquisition of English auxiliary verbs; if there is L1 transfer, to estimate how long the effect lasts; and finally, to look for other factors on the acquisition that cannot be traced to the L1, i.e. developmental factors that may be uniform across all children acquiring L2 English. Data from a group of children selected from the longitudinal study are used, and the children are split into the same L1 groups as in the previous chapter, namely Chinese, Hindi, Spanish, and Arabic. Several predictions for child L2 learners are tested. The predictions concern accuracy in obligatory contexts for auxiliaries BE and DO in declarative and interrogative sentences, error patterns in auxiliary use, and whether the findings provide evidence for transfer of syntactic structure from L1.

This chapter is organized as follows: the next section briefly summarizes findings of previous studies on L2 acquisition of morphosyntax that motivated the present study (a detailed review was given in Chapter 3, Section 3.4.3) and reminds the reader about possibilities for transfer of L1 features into L2 English in terms of the theoretical models (discussed in Chapter 3, Section 3.6). Section 6.2 summarizes the methodology used in the auxiliary elicitation task and the data are coded. The results are separated into discussions of accuracy in declaratives (Section 6.3.1) and questions (Section 6.3.2). Section 6.4 summarizes the findings with regard to the role of L1 in children's acquisition of the auxiliary system.

6.1.1 Background and aims of the present study

As discussed earlier in Chapter 3 (Section 3.4.3), English verb morphology is one of the areas that have been substantially researched in L1 acquisition as well as in adult and child L2 acquisition. However, there were no studies that systematically analyzed the acquisition of auxiliaries BE and DO in child L2 acquisition, since most child L2 studies concentrated on the acquisition of tense and agreement morphology or the semantics of aspect. Thus, there was little discussion of how the syntax of auxiliary verb constructions develops. Bringing together the findings in previous work on child L2, we can see several trends, namely that omissions of BE are frequent in child L2 acquisition, the sentence structure plays a role, as declaratives with BE are acquired earlier than questions, and that questions with DO appear to be the most difficult structure. Furthermore, the role of L1 in child L2 acquisition is an area of investigation that still has not generated a clear conclusion in previous research (Chapter 2, Section 2.2.1). This chapter will discuss the role of L1 in L2 children's acquisition of three morphosyntactic structures: declarative and interrogative sentences with auxiliary BE and interrogative sentences with auxiliary DO. The general research questions that will be addressed are parallel to the questions in Chapter 5 on the acquisition of articles: (1) Is there evidence for L1 influence at the early stage?; (2) Are there developmental trends interacting with L2 trends?; and (3) Does the L1 influence change with is the amount of exposure to English? Specific predictions for the participants in the present study, tied to the contrasts represented in their L1s, will be recapitulated in the following section.

6.1.2 Contrasting English and children's L1s

The L1 backgrounds of the children represent an interesting mix of typologically different languages with structures both similar to the English auxiliary constructions and different from them. Aspectual constructions with auxiliary BE in English and its equivalent ESTAR in Spanish have been analyzed similarly in previous studies (See Chapter 3, Sections 3.4 and 3.5.2), which argued that formally these languages can be analyzed as instantiating the abstract category of

Aspect. Structural analyses of word order changes in questions are also similar in Spanish and English. For English and Spanish, it was argued that the word order in questions can be accounted for in terms of the movement of the inflected verb from its position in the Aspect Phrase to the Complementizer Phrase. Since Arabic and Hindi have auxiliaries that are the spell-out of features [perfective] and [progressive], the formal account of the structure of declarative sentences can be applied to these languages as well, with Inflection and Aspect phrases being projected in the sentence structure. However, in these languages overt verb movement does not take place in questions. Arabic is different from English, Spanish, and Hindi because even though it has an auxiliary verb similar to BE that forms part of an analytic progressive construction, it exists only in the past tense. Thus, Arabic is formally analyzed as similar to the three languages already mentioned, but with the Aspect projection unfilled in the present tense.

In contrast with Spanish, Hindi, and Arabic, tense in Chinese is not marked morphologically on verbs. Rather, it is marked lexically using time adverbs. In formal syntactic analyses of tense, it has been argued that there is no syntactic distinction of finiteness/non-finiteness in Chinese languages, and thus no projection of Tense or Inflection. It was argued that since aspectual features are grammaticalized as verb affixes, they need to be accounted for in terms of a functional category Aspect in the sentence structure. Under such an account, functional projections in clauses in Mandarin (and presumably also in Cantonese) would include Aspect Phrase and Verb Phrase, but no Inflectional Phrase.

In the group of background languages represented in this study, Spanish is the only language with subject-verb inversion in questions. However, in Spanish the inversion rule applies both to auxiliaries and to main verbs, whereas in English it applies only to auxiliaries. Questions with uninverted subject-verb order appear to be more acceptable in Spanish than in English, but out of all L1s in the study, Spanish is the only language that employs the syntactic operation of subject-verb inversion in questions.

To summarize, Arabic, Hindi, and Spanish have a full range of progressive constructions with an auxiliary similar to the English BE inflected for person and

number. In Arabic, however, the progressive construction is used only in the past tense. Only Spanish has subject-verb inversion in questions (BE and main Verbs). Finally, none of the languages have an auxiliary similar to DO in yes/no questions.

6.1.3 Acquisition models to be considered

In order to test whether there are L1-related differences in the way children acquire auxiliaries BE and DO, their utterances will be analyzed in two ways. The children's accuracy with auxiliaries, i.e. the percentages of suppliance of correct forms in obligatory contexts, will be analyzed. Mean percent correct scores will be compared across L1 groups. In line with the analyses in the previous chapter, the Full Transfer/Full Access model of L2 acquisition will be applied in order to evaluate the data from a generative point of view. FT/FA model predicts that the knowledge of abstract grammatical categories such as Tense and Agreement is transferred from a L2 learner's L1. If a certain category is not instantiated in L1, it will be supplied from the universal inventory of categories that learners have access to. Children whose L1s do not have these abstract categories initially will not be aware that it is obligatory and will omit auxiliaries. Gradually they will reconstruct their mental representation of the clause structure.

As for syntactic inversion in questions, we can expect L2 children to acquire the operation of inversion successfully. But if we consider the possibility of transfer of L1 word order, learners whose L1s employ subject-auxiliary inversion are expected to have an advantage at the initial stage of acquisition. However, different predictions can be made for the acquisition of *do*-support in questions, since the formation of questions with DO auxiliary is a structure that is specific to English. If L1 transfer cannot facilitate L2 acquisition of *do*-support, all L2 children are expected to learn the DO-insertion and movement of auxiliary from input. Predictions can be summarized as follows:

(1) Full Transfer/Full Access predictions

a. Children whose L1s project Tense and Aspect in the sentence structure are expected to transfer this knowledge to L2 English and be more accurate with the provision of BE than the children whose L1s do not instantiate these categories.

b. All children are expected to have difficulty with DO-questions because it is an idiosyncratic structure of English, resulting in lower accuracy with DO than with BE.

In the discussion of generative approaches to L2 acquisition in Chapter 2 (Section 2.2) the FT/FA account was compared with the feature-assembly model. Feature assembly, like FT/FA, is based on the assumption about the presence of abstract categories in L2 initial state, but unlike FT/FA, it allows us to make more specific predictions for L1 transfer based on its morphosyntactic features. Under a feature-assembly account, L2 learners start off with a fully assembled set of L1 lexical items and grammatical categories. Thus, the Spanish-, Arabic-, and Hindi-speaking learners are expected to transfer the knowledge of the functional categories Tense, Agreement, and Aspect to their L2 English. The difference among the groups is that the Spanish speakers tense- and aspect-related features are mapped onto vocabulary items (auxiliary verbs) in the same way as in English for progressive constructions with BE. In contrast, in Arabic these features are mapped onto overt auxiliaries only in the past tense, whereas in Hindi Tense and Aspect are mapped onto two separate auxiliaries. All these languages are contrasted with Mandarin and Cantonese Chinese, in which the abstract category of Tense is not projected, and Aspect features are mapped onto sentential particles (See Table 3.1 in Section 3.6 of Chapter 3). With respect to learners' accuracy, the feature assembly approach allows us to make the predictions as follows:

(2) Feature assembly predictions

a. For auxiliary BE, the Spanish group will be more accurate than the Arabic and Hindi groups, which in turn will be more accurate than the Chinese group

b. All four L1 groups are expected to demonstrate low accuracy with DO

Furthermore, both the FT/FA and the feature assembly account predict L1 effects in the acquisition of syntactic inversion: If the Spanish children transfer the knowledge of inversion from L1, there should be no difference between their accuracy with declaratives and questions with BE, because inversion of BE employs the same syntactic operation in English and Spanish. If the Arabic-, Chinese-, and Hindi- speaking children transfer the syntactic structures from their L1, we expect to find their knowledge of inversion to be initially faulty, since their first languages to not require any changes in word order for question formation.

6.2 Method

As discussed in Chapter 4 (Section 4.2.3), children's production of grammatical morphemes in elicited speech will be assessed using the grammatical probes of the Test of Early Grammatical Impairment (TEGI, Rice & Wexler, 2001). The following sections describe the groups of participants selected for the present study and discuss the experimental procedure and the procedure used for coding the data.

6.2.1 Participants

Recall that the analysis of children's acquisition of articles used data both from the longitudinal and from the cross-sectional study. The present analysis of the children's acquisition of auxiliary verbs focuses only on the cross-sectional study because this study has larger numbers of participants in all L1 groups. For the analysis of L1 transfer effects, the L1 groups will be the same as the four groups in the cross-sectional study of the children's acquisition of articles (Chapter 5), namely Arabic, Spanish, Chinese, and Hindi. These languages

represent interesting contrasts with respect to their auxiliary systems, discussed earlier in Chapter 3 (Section 3.5). Recall that in Chapter 5, L1 effects in the domain of articles were observed only at the early stage of the longitudinal study (after 2 to 18 months of exposure to English). Similarly, in the cross-sectional study, L1 effects were found in the children with no more than 18 months of exposure. Thus, if this is indeed the time frame when we can see significant L1 effects in the acquisition of L2 morphosyntax, we should expect L1s to influence the use of auxiliaries at this stage as well. Information about these children was summarized in Table 4.4 (Chapter 4, Section 4.1.2). In order to form the early-stage groups, only participants with less than 21 months of exposure were selected. While the maximum MOE for participants was limited to 18 months in the study of articles in the previous chapter, the upper limit on MOE had to be increased to 21 for the present analysis because of the difficulty of the elicitation task for children with little exposure to English, which resulted in insufficient scorable responses. For the same reason the number of participants was increased to 12 in each of the four L1 groups, as opposed to 10 in the previous analysis.

6.2.2 Materials

As discussed in Chapter 4 (Section 4.2), data collection was a set of measures for assessment of children's L2 development including recordings of a spontaneous speech sample, as well as elicited speech. The grammatical probes of TEGI were chosen as the best assessment of the mastery of auxiliaries in elicited speech. The scenario for the elicitation of auxiliaries BE and DO was described in detail in Chapter 4 (Section 4.2.3). Example (3) gives the three main target structures elicited in the test.

(3) a. BE in declaratives:
The bug is jumping.
b. BE in questions:
Are the moon guys crying?
c. DO in yes/no questions in TEGI

Does the bug need a tissue?

In total, the BE/DO probe comprised 11 attempts of DO-questions, 6 attempts for BE-statements, and 6 attempts for BE-questions.

6.2.3 Coding procedure

A detailed description of the coding procedure and illustrations were given in Chapter 4 (Section 4.2.3). Recall that three structures were included in the analysis: declaratives with BE, questions with BE, and questions with DO. All obligatory contexts for the elicitation of auxiliary BE were third person (singular or plural) present tense. Utterances with no subject (e.g. *Are jumping*) or with an auxiliary other than BE (e.g. *Bears can jump*) were excluded. Elliptical utterances that consisted of just the participle form of the main verb were excluded because such utterances could not be unambiguously judged as marked or not marked for finiteness, or it was not clear what the syntactic subject of the sentence was. Uses of BE were coded as correct in obligatory contexts when the auxiliary was appropriately inflected for person, number and tense and accompanied by the participle form of main verb. The errors in BE contexts were the omission errors (utterances with only the main verb present in participle form: *They snoring*) and commission errors (utterances with a wrong form of BE: *Bears is snoring*). In questions with BE, errors also included double marking, when the question was formed with the auxiliary BE used twice, at the beginning of the sentence and after the subject (e.g. *Is the bug is laughing?*). Thus, all scorable responses were full progressive constructions with the subject, auxiliary and participle present. In question elicitation, children sometimes produced questions with a correct form of BE, but in an uninverted declarative form. Such utterances were coded as uninverted questions only when they were pronounced with what was clearly a rising question intonation. They were not included for the accuracy counts, but they will be discussed separately in the analysis of error types.

Responses with DO were scored as correct when the auxiliary was used with correct tense and person agreement, and the main verb was uninflected.

Commission (wrong form) errors were uses of DO with wrong agreement (DO for DOES or DOES for DO), wrong tense (DID for DO), or wrong auxiliary (e.g. BE for DOES: *Is he like juice?*). Double marking errors refer to utterances with DOES in which the main verb was inappropriately inflected with 3sg *–s* (e.g. *Does he likes juice?*). Individual scores for various error types were very low because there were 6 elicitation contexts for each of the two BE structures and 10 contexts for DO questions. Since some of the children's responses were unscorable, individual tokens of each error were often less than 3. Thus, percentages of error types were not calculated. The analyses below are based only on the children's percentages of accurate responses.

6.3 Results

This section discusses the results of the analysis of the children's responses in the BE/DO probe. Separating the children into four groups according to the typology of their L1, I analyze auxiliary provision and omission in obligatory contexts. The main focus of the analysis is on the subset of children who are in the early stage of acquiring English (less than two years of exposure) because the previous study of the children's acquisition of the article system revealed that L1-based differences between groups became weak or disappeared after approximately 1.5 years of exposure to English. The Chinese, Hindi, Spanish, and Arabic groups are contrasted with respect to their accuracy with auxiliary suppliance and with respect to error types. Statistical analyses of the children's performance are performed in order to ascertain the effect of the L1 typology and the amount of exposure to English.

6.3.1 Analyses of accuracy

The general research questions asked for the analyses of accuracy with auxiliaries at the early stage of L2 acquisition were the following: Are the rates of accurate provision of auxiliaries affected by L1 background in the four L1 groups? Are accuracy rates influenced by sentence type (declarative or interrogative)?

Declarative statements and questions were analyzed separately because it was important to assess children's ability to produce inverted structures, which are presumably more complex syntactically, and also because the mastery of syntactic inversion might be influenced by L1.

Accuracy in declaratives with BE

Children's accuracy in declarative contexts was calculated as a percentage of all scorable responses that included a form of the auxiliary correctly inflected for number and tense (recall that all the target responses in the test had to be in third person singular, present tense form). Individual accuracy scores that were used to calculate the mean are given in Appendix 4. The resulting mean percentages of correct use for each L1 group are shown in Figure 6.1.

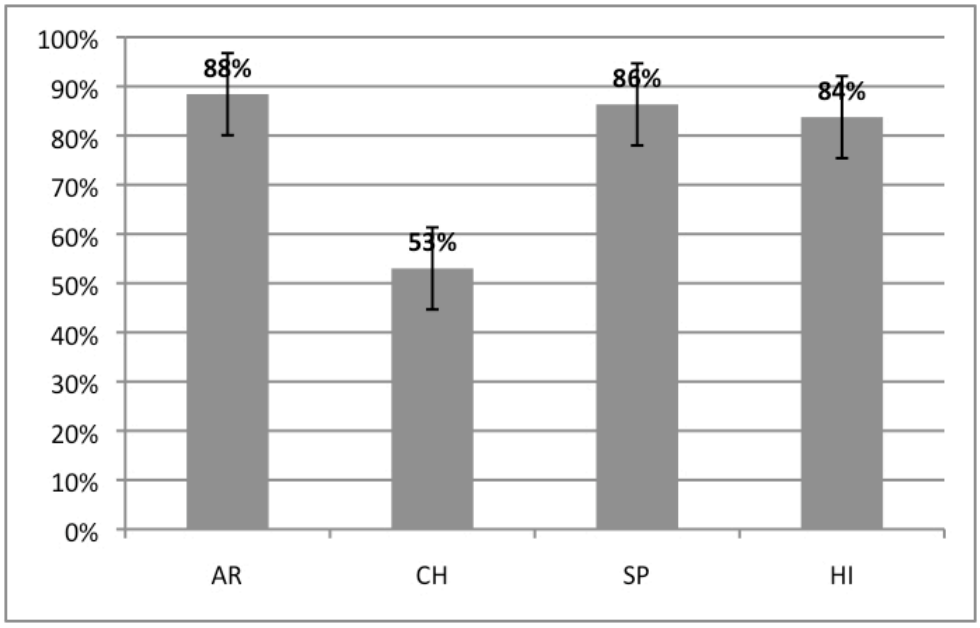

Figure 6.1. Mean percent correct scores for auxiliary BE in declaratives for the Arabic (AR), Mandarin/Cantonese Chinese (CH), Spanish (SP) and Hindi/Urdu/Punjabi (HI) L1 groups

The obvious pattern that is revealed in the comparison of the children's accuracy scores is the contrast between the L1 Chinese group and the rest of the groups. To determine whether the effect of L1 was significant in these data I conducted a

one-way between-subjects ANOVA that yielded a significant effect of L1 on accuracy with BE ($F(3,47) = 4.82$, p=.005, eta squared = .25). The apparent facilitation effect in the Arabic, Spanish and Hindi groups was further explored in post-hoc comparisons. Tukey honestly significant difference (HSD) pairwise comparisons on the group means for BE in declaratives revealed that the Chinese L1 group mean accuracy score was significantly lower than the scores of the Arabic, Spanish and Hindi L1 groups (mean differences between the Chinese and other three groups were 35%, 33%, and 31% respectively, p = .010, .017, and .032), whereas there was no difference among the three groups.

Accuracy in questions with BE

Recall that correct responses in the elicitation task for BE-questions had to include a form of BE correctly inflected for person number and tense, a syntactic subject, and a participle form of the main verb. Sentences with question intonation but uninverted subject-auxiliary order were not considered for this analysis. The accuracy data for the four L1 groups are provided in Figure 6.2. Individual accuracy scores are given in Appendix 4.

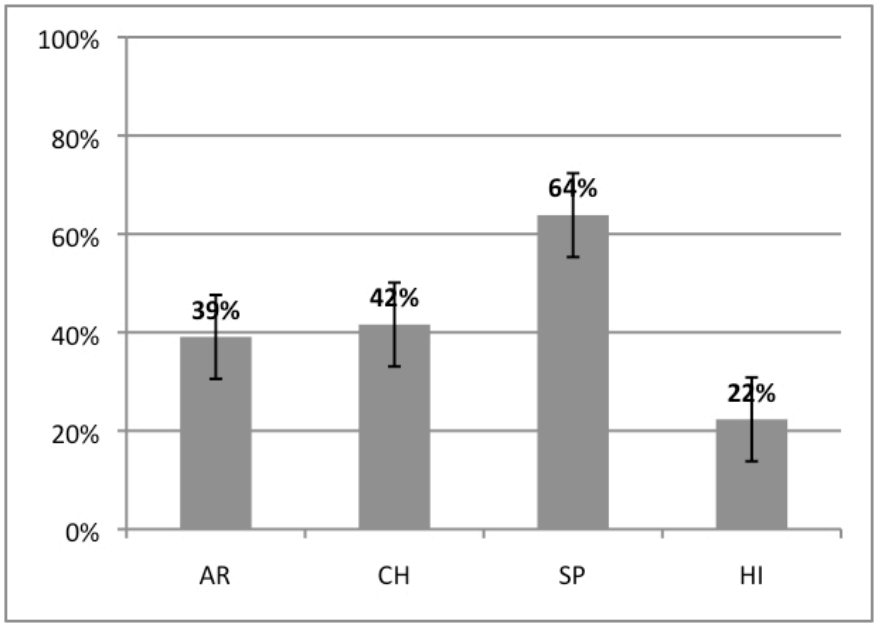

Figure 6.2. Mean percent correct scores for auxiliary BE in questions

Several trends can be observed in the data. First of all, the general trend is that accuracy scores are much lower in questions than in declaratives discussed in the previous section. Secondly, with respect to L1-based differences, two groups stand out, namely the Spanish group with the highest mean accuracy score of 64%, and the Hindi group with the lowest score of 22%. A one-way ANOVA confirmed that the effect of L1 on the performance in questions was significant ($F(3,47) = 2.97$, $p = .042$). However, the effect of L1 in these contexts was very small (partial eta squared = .18). Post-hoc Tukey HSD comparisons on accuracy scores with BE questions revealed only one significant difference between the Spanish and the Hindi L1 group (mean difference 41%, $p = .024$), and there were no differences among other pairs.

Accuracy in questions with DO

The third type of structure elicited in the TEGI was questions with DO. Comparing the children's performance for this type of question, no differences were predicted among the L1 groups, as all L1 groups were expected to have more difficulty with the auxiliary DO because it is a language-specific feature of English. Recall that children's accuracy in DO contexts was calculated as a percentage of all obligatory contexts in which children produced a form of auxiliary DO correctly inflected for person, number, and tense, and correctly inverted. . Individual accuracy scores are given in Appendix 4. The resulting mean percentages of correct use for each L1 group are presented in Figure 6.3.

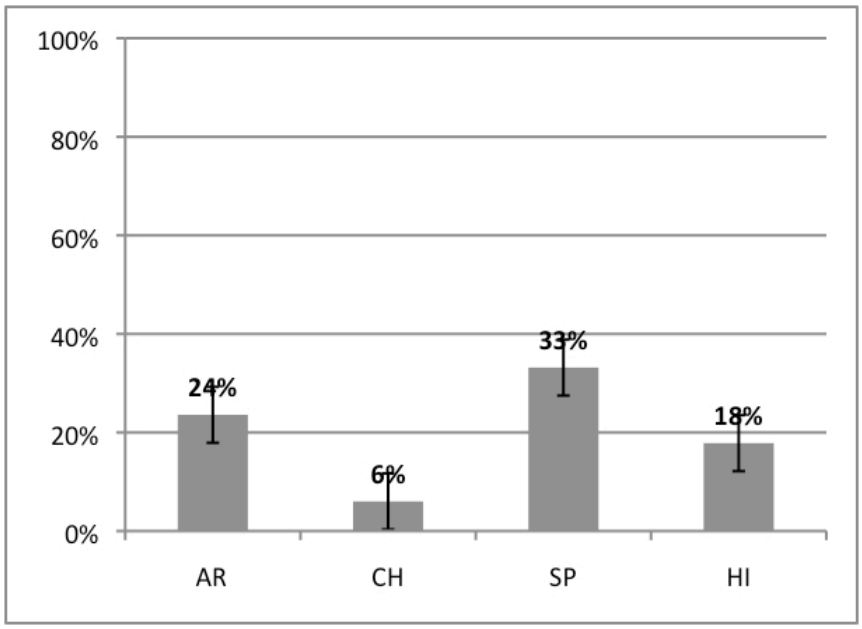

Figure 6.3. Mean percent correct scores for auxiliary DO in questions

The performance of all groups in DO contexts was clearly less accurate than their performance in both declarative and interrogative contexts for BE. Interestingly, there appears to be substantial variation in accuracy scores across L1 groups. However, a one-way ANOVA did not yield a significant effect of L1 for these children's accuracy with DO ($F(3,47) = 2.48$, $p = .072$, partial eta squared = .14). The absence of a significant effect of L1 could partly be due to missing data points for a number of children in some groups, which were the result of a high number of unscorable responses. In the Chinese and Spanish L1 groups, 9 children out of 12 contributed to the score, and in the Chinese L1 group, only four children produced any correct responses. Unscorable responses in DO contexts were declarative sentences (e.g. *Maybe the bugs like milk*) or a repetition of the experimenter's prompt (e.g. *I wonder if the bugs like milk*). Compared to the number of children who produced correct responses in the Arabic, Spanish, and Hindi groups (9, 9, and 7 respectively), this might be interpreted as an indication that L1 typology influenced the children's performance in this task as well, since the Chinese L1 group appeared to have the most difficulty producing questions with DO, while the Spanish group had the least difficulty. While this is only a

speculation, these trends might prove significant with larger numbers of participants.

Finally, comparing the children's performance across types of sentence structure rather than across L1s, it is clear that some structures were more problematic than others. In order to determine whether the effect of structure type was significant, a two-way mixed ANOVA was conducted with L1 group as a between-subjects factor (four levels) and structure type as the within-subjects factor (three levels). Structure type effect was significant ($F(2,43) = 8.616$, $p < .001$). The ANOVA also yielded a significant interaction of L1 and structure type ($F(6,86) = 2.385$, $p = .035$). Post-hoc paired samples *t*-tests confirmed the overall accuracy pattern, with significant differences between all pairs of structures: BE declaratives and BE questions ($t = 6.16$, $p < .000$), BE declaratives and DO questions ($t = 13.24$, $p < .001$), and BE and DO questions ($t = -4.42$, $p = .001$). For all the children taken together, the accuracy was the highest in declarative sentences with BE, and it was the lowest in questions with DO.

6.3.2 Uninverted questions with BE and DO

The following predictions were made regarding the acquisition of subject-auxiliary inversion: If the Spanish children transfer the knowledge of inversion from L1, there should be no difference between their accuracy with declaratives and questions with BE, because inversion of BE employs the same syntactic operation in English and Spanish. If the Arabic-, Chinese-, and Hindi- speaking children transfer the syntactic structures from their L1, we expect to find their knowledge of inversion to be initially faulty, since their first languages to not require any changes in word order for question formation. In contexts for DO, all children were expected to lack *do*-support and inversion. Examples in (4) illustrate uninverted questions in contexts where DO was elicited, and sentences in (5) are examples of uninverted BE questions.

(4) The kitty wants milk? (75, L1 Arabic, 5;00, 5 MOE)
The moon guys like orange water? (05, L1 Spanish, 5;05, 13 MOE)

(5) The moon guys are resting? (04, L1 Spanish, 5;04, 4 MOE)
He's crying? (04, L1 Spanish)

Mean proportions of uninverted questions out of all responses in question elicitations were calculated for each L1 group. Unlike in the previous analyses, the denominator in the proportions was the total number of question attempts, including uninverted questions. The resulting group means for BE and DO contexts are given in Figure 6.4.

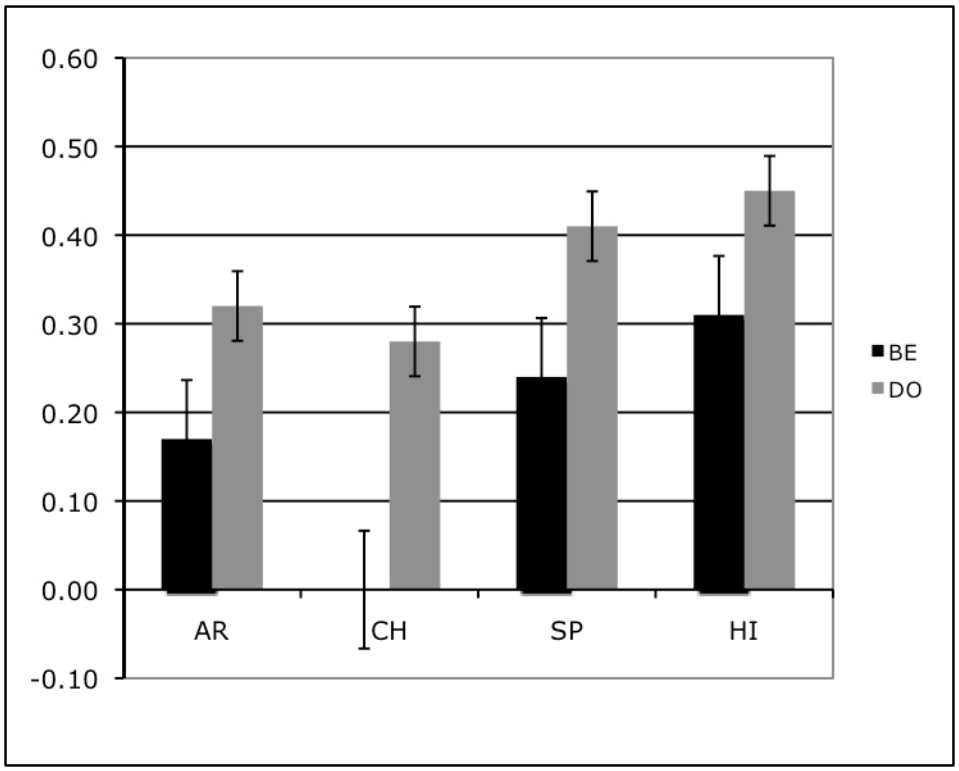

Figure 6.4. Mean proportions of uninverted structures out of all responses for questions with BE and DO

An interesting finding was the absence of uninverted BE-questions in the Chinese L1 group, and similar proportions of uninverted questions in the three other groups. This finding is not surprising, however. It was already determined in the analysis of children's accuracy with BE that the Chinese L1 group had the lowest rate of suppliance in declarative structures, while other groups' mean accuracies

ranged from 84% to 88% correct. As for the rates of uninverted questions in DO contexts, the analysis revealed similar proportions of uninverted questions in all L1 groups (ranging from 27% to 45%), indicating that all children similarly resorted to using declarative sentence structure with question intonation instead of using *do*-support and subject-auxiliary inversion.

6.4. Summary

Are there L1 effects in the acquisition of auxiliaries?

Recall that the Full Transfer/Full Access model predicted that children whose L1s project Tense and Aspect in the sentence structure would transfer this knowledge to L2 English and be more accurate with the provision of BE. The feature assembly account predicted more specific differences among the groups, since the grouping of tense and agreement features on auxiliary verbs is the most English-like in Spanish, and less so in Arabic and Hindi, whereas the Chinese languages completely lack lexical items parallel to the English auxiliary verbs. Thus, the Spanish-speakers were predicted to be the most accurate, the Hindi- and Arabic-speakers were predicted to be less accurate, with the Chinese-speakers being the least accurate of all the groups. With respect to accuracy with auxiliary BE, the only significant L1 effect apparent in the data was the difference in accurate suppliance of BE in declarative sentences. The significant L1 effect on accuracy was due to the low suppliance of the auxiliary in the Chinese L1 group, while the other three groups were equally producing correct declaratives with BE. There were no significant differences in the accuracy among the Arabic, Spanish, and Hindi L1 groups. These findings support the predictions of the FT/FA model, indicating that the presence of the abstract categories of Tense and Agreement facilitated acquisition, rather than more specific similarities in the grouping of features into lexical items predicted by the feature assembly account.

Both FT/FA and feature assembly models predicted that all children would have difficulty with DO-questions because it is an idiosyncratic structure of English, resulting in lower accuracy with DO than with BE. This prediction was

borne out in the data, as the statistical tests revealed no significant differences among mean accuracy rates with DO. Interestingly, there appeared to be substantial, albeit insignificant, variation in accuracy scores across L1 groups. It might be the case that the knowledge of L1 Spanish facilitated the acquisition of the new structure due to the transfer of abstract knowledge of inversion, but the data are insufficient to establish a significant group trend.

L1 transfer and syntactic inversion

Both transfer models considered in the study predicted that since the Spanish children should be able to transfer the knowledge of syntactic inversion from L1 because question formation employs the same syntactic operation in English and Spanish. Since the Arabic-, Chinese-, and Hindi- speaking children are also expected to transfer the syntactic structures from their L1, their knowledge of inversion was expected to be initially faulty, since these languages to not require any changes in word order for question formation.

Statistical tests revealed that uninverted questions constituted similar proportions of question attempts in the Spanish, Arabic, and Hindi L1 groups (recall that all instances of correct questions analyzed in Figures 6.2 and 6.3 included a correct auxiliary form *and* were also inverted). It might seem surprising that the Spanish-speakers used a high proportion of uninverted questions with BE (24%), since Spanish has the syntactic operation of inversion, similar to English. However, it was mentioned in the language descriptions in Section 1.3 that the use of declarative structure with rising intonation to mark questions appears to be more acceptable in Spanish than in English. What was not surprising was the absence of uninverted BE questions in the Chinese group. Since the Chinese languages altogether lack an auxiliary similar to BE, the speakers of these languages could not employ the declarative progressive construction with question intonation as did the speakers of other languages in the study.

The role of structure type

Statistical tests revealed that, across all L1 groups, mean accuracy with auxiliaries was the highest in declarative sentences with BE, and it was the lowest in questions with DO. This finding again points to the similarity between L1 and child L2 acquisition mentioned in the review of previous studies in Section 1.1. Both in L1 and L2 acquisition of English, children appear to have more difficulty producing questions than declaratives with BE. Previous usage-based studies of naturalistic (Rowland, 2007) and elicited data (Theakston & Rowland, 2009) report that monolingual English children perform more poorly on questions than declaratives. This developmental trend appeared to be stronger than L1 based trends in the study, suggesting that the common difficulty with questions might be due to the inherent difficulty of the syntactic operation of inversion (contra Santelmann et al.'s (2002) claim that the knowledge of inversion is readily available to all learners from UG), or it could also be due to input frequencies of auxiliaries as parts of question constructions (see discussion of Lieven, Rowland, Theakston and colleagues' studies in Section 3.4.3 in Chapter 3). This possibility will be considered in more detail as part of the analyses of the role of input in Chapters 7 and 8.

CHAPTER 7. CONSTRUCTIONS IN CHILD L2 ENGLISH: THE ARTICLE SYSTEM

In the following sections, children's suppliance of the definite and indefinite articles will be analyzed from a usage-based perspective. It was already mentioned in the discussion of the findings in Chapter 5 (Section 5.5) that input frequency could be one of the factors that make the article *the* 'easier' to learn for the L2 children, and that they were more accurate using articles in certain syntactic constructions. This chapter will follow up on these observations and focus on the children's suppliance of articles in indefinite contexts, where most errors were found. While the analyses in Chapters 5 and 6 investigated the issue of L1 transfer in child L2 acquisition, the analyses in the present chapter will pertain to the theoretical issues of the role of frequency and emergent productivity in L2 acquisition (introduced in Section 1.3 of Chapter 1 and in Chapter 2). The analyses in this chapter will compare the frequency of *the* and *a* in the input compare and in the children's speech, investigate the distribution of articles with respect to construction type in the input and in the children's speech, and follow the development of children's productivity with articles over time.

7.1. Introduction

Recall that the children in the longitudinal study described in Chapter 5 were divided into two groups based on L1: a [+article] and a [–article] group. While the study found evidence for L1 effects in the children's acquisition of articles, it also found that both the [+article] and the [-article] L1 groups were more accurate with *the* in definite contexts than with *a* in indefinite contexts (Chapter 5, Section 5.3). This difference was present at all stages of acquisition, and across learners with different L1 backgrounds, and thus, was a robust pattern. Another finding was that *the* misuse in indefinite contexts clearly was the dominant error type for both the [+article] and [–article] groups. In addition, there was a pattern in the children's article use that that did not seem to depend on the children's L1 and it did not follow from the theoretical accounts: in both the

longitudinal and the cross-sectional studies, the children appeared to supply the indefinite article correctly in some syntactic positions more frequently than in others. In utterances with the existential *there* (e.g. *once upon a time there was **a** rabbit*), the indefinite article was never omitted or substituted with *the*, while *the* overuse was very common in all other indefinite contexts. This pattern was present in children from all L1 backgrounds. In addition, the children often used an indefinite article accurately when the target noun was introduced as an object, e.g. *the dog was building **a** castle*. In the subject position, nouns were often introduced erroneously with a definite article. Precisely these differences will be explored further in this chapter. First consider examples of stories in which one of the nouns was introduced with the existential *there*, and the other one was not:

(1) One day there was **a elephant girl** that have some ball. She said to **the giraffe** that they can play in the water with the ball.
(20, L1 Spanish, 5;01, 4 MOE)[8]

(2) There was **a person** seeing another girl. They were going for picnic. And they saw **the doctor**.
(61, L1 Urdu, 5;02, 11 MOE)

(3) Once upon a time there was **a girl rabbit**. And sh… he was pulling her wagon.
Then **the brother rabbit** came running after the girl.
(88, L1 Punjabi, 5;11, 17 MOE)

In the three examples, indefinite articles are supplied correctly when the indefinite nouns *an elephant girl*, *a person*, *a girl rabbit* follow the words *there was*. Within the same story, definite articles are supplied incorrectly with the nouns that require an indefinite article but are not introduced with *there was* (i.e. *the giraffe*, *the doctor*, *the brother rabbit*). In fact, the children never overused the definite

[8] In the parentheses, I indicate the child's name code, first language, age, and months of exposure to English (MOE) if it was the cross-sectional study or round of data collection (R) if it was the longitudinal study. In the cross-sectional study, name codes consisted of 2-3 digits, and in the longitudinal study they consisted of four letters.

article in the existential *there is/was* context. This finding indicates that it was not the noun that influenced the article choice (i.e. it was not the case that phrases *a girl* or *the brother* were memorized as wholes), but rather the words preceding the NP. The influence of context on accuracy with articles needs to be investigated further.

High accuracy with article suppliance in contexts such as *there* construction was reported earlier in adult L2 learners of English (Lardiere, 2005; Trenkic 2007; White 2003, 2007). For instance, White (2003), in a study of a Turkish speaker whose proficiency in English was advanced, reports no *the* overuse in the existential *there* context, even though the subject did make errors in article suppliance. Lardiere (2005), in her study of a steady-state L2 speaker whose L1 was Chinese, similarly reports no *the* overuse in this particular context. White (2007) discussed the possibility suggested by Trenkic (2007) that 'There is a ….' can be memorized as a chunk on the basis of frequency in the input, but concluded that it is unlikely to be the whole explanation. Evidence used to counter the frequency account included the variety of NPs that the L2 learners produce (e.g. *there are some*…, *there are a few*…), as well as the fact that many of them exhibited appropriate plural agreement (*there is a*... vs. *there are*…), suggesting productivity. Using data from the longitudinal study, progress in the children's accuracy with indefinite nouns will be analyzed over time in order to establish whether, at the early stages, the children used the indefinite article exclusively as a part of the existential *there* construction.

Another construction-dependent use of the indefinite article was observed with indefinite object nouns, which were accompanied by an indefinite article more often than indefinite subject nouns. This asymmetry is illustrated in (2) and (3) with the correct use of the indefinite article with the object noun. Another typical contrast of correct and incorrect article suppliance is illustrated in (4), where the indefinite article is supplied correctly only with the noun in the object position:

(4) **The girl rabbit** was making **a castle**.
And then **the boy rabbit** came.
(88, L1 Punjabi, 5;11, 17 MOE)

It was the case for some children that, in each story, the only correct uses of the indefinite article were found with nouns in object positions. The asymmetry in the suppliance of target articles in subject and object contexts was observed in other learner populations in the previous research. For instance, Chondrogianni (2008) reported that Turkish-speaking children learning L2 Greek omitted more articles in subject than in object position. This pattern, according to Chondrogianni, is consistent with the findings in L1 Greek (Marinis, 2003). However, it could be due to the properties of Greek articles, and further investigation of syntactic contexts of article use in L1 acquisition of English is needed in order to establish whether subject-object asymmetry is a developmental pattern in the acquisition of the article system.

The role of syntactic constructions and input frequency in the distribution of definite and indefinite articles in L2 children will be investigated in terms of the usage-based model in the following sections, with the aim of answering the following research questions:

(1) How does the frequency of *the* and *a* in the input compare with that in the children's speech? Can the frequency of articles in the input account for the order of acquisition (*the* before *a*)?

(2) How are articles distributed with respect to construction type in the input and in the children's output?

(3) Is article use influenced by the collocates of articles, e.g. the preceding verb in the case of direct object NPs? In an analysis of article use over time, is there evidence for construction-specific use of articles at the early stages of acquisition, with gradual increase in the variety of contexts in which articles are used correctly?

7.2 Method

7.2.1 Participants

The analysis in this section will use longitudinal data from the same corpus that was used as the source of data in Chapter 5. Recall that the corpus included data from 16 participants (Table 4.1 in Section 4.1). For a more detailed analysis, nine children were selected from the larger corpus. Table 7.1 gives the background information about these children. After excluding participants whose data collection was interrupted or who did not take part in each round of the study, the number of participants in each L1 group was limited to three. The objective of selecting a small subset of participants was to pursue more detailed case-study analyses of each child's development over time. For the same reason, the cross-sectional data were not used for the analyses in Chapters 7 and 8. In these chapters, the focus is on the individual development of the productivity of various constructions in a child rather than generalizations over large groups of children.

Table 7.1

Children's L1, age and months of exposure (MOE) to English at Round 1 of the study

L1 and name	Age	MOE
Arabic-LLKC	4;10	11
Arabic-TRRK	4;02	8
Arabic-YSSF	4;11	9
Spanish-DVDC	6;03	8
Spanish-SBST	5;01	15
Spanish-SMNS	5;07	6
Mandarin-DNNC	5;04	9
Mandarin-MRSS	5;00	4
Mandarin-JNNH	5;11	18
mean	5;03	10

Recall that there five rounds of data collection in the study and they were six months apart.

7.2.2 Materials

Children's narratives

For the present study of the children's use of articles, the corpus of narratives will again be used as the source of data (see Chapter 4, Section 4.2.2). Recall that in the process of scoring elicited stories, only the use of articles was considered. All other ways to refer to characters, such as personal and deictic pronouns and proper names were excluded. The analysis was limited to the use of articles with singular countable nouns because these nouns required an article in all contexts. Each noun included in the analysis represented one of the two kinds of obligatory contexts for articles, indefinite or definite, and there were three possible codes for article use in each context:

Indefinite: - correct (*a*+noun, e.g. *He had a plane*)
- substitution (*the*+noun, e.g. *He had the plane*)
- omission (bare noun, e.g. *He had plane*)

Definite: - correct (*the*+noun, e.g. *The plane fell into the pool*)
- substitution (*a*+noun, e.g. *A plane fell into the pool*)
- omission (bare noun, e.g. *Plane fell into the pool*)

See Section 4.2.2 for a detailed discussion of the coding procedure and examples.

Adult speech: Online corpora of American English

The L2 input that the children received was mostly English spoken at school. Since it was not feasible for this study to document the English that the children in Edmonton heard from adults and other children, it was necessary to find a corpus of spoken English to use as a model of the speech that the children were exposed to. The Corpus of Contemporary American English (COCA: Davies, 2008) was used in order to assess the distribution of definite and indefinite articles in adult spoken English. The corpus was searched using the tools and interface developed by Mark Davies (http://www.americancorpus.org). The size of the spoken subcorpus of COCA is approximately 81 million words, which makes it the biggest corpus of spoken English. Even though the children in the study had

exposure to Canadian English, there is no principled reason to think the relative distribution of definite and indefinite articles would be substantially different in American and Canadian varieties of English.

7.3 Results

The analyses in this section will test the following predictions of the usage-based model:

1. If the frequency of articles in the input is reflected in the order of acquisition, *the* will be acquired before *a*;

2. The distribution of articles with respect to construction type in children's speech will be similar to that in the input:

- Article use will be influenced by collocates of articles, e.g. the preceding verb in the case of direct object NPs.

- The use of articles at the early stages of acquisition will be construction-specific, with gradual increase in the variety of contexts in which articles are used correctly.

7.3.1 Distribution of articles in adult speech

In the usage-based model of language acquisition, the order in which children acquire articles is expected to reflect their relative frequencies in the language children hear. It was already mentioned in the discussion in Chapter 5 that the children's accuracy with *the* could be partly due to its high frequency in English. Using the [at*] part-of-speech tag for 'article', frequencies of *a*, *an*, and *the* were retrieved from the spoken sub-corpus of COCA. The definite article *the* occurred 3,860,303 times in the corpus, and the indefinite article *a(n)* occurred 2,008,757 times. The definite article was the most frequent word in the COCA, according to the frequency list, while *a* occupied the 5th place. While both articles are among the most frequent words in English, *the* is almost twice as frequent as *a*, which was reflected in the children's order of acquisition. It was demonstrated in Chapter 5 (Section 5.3) that the children in this study started using *the* before *a* and were more accurate in supplying *the* in obligatory contexts throughout the

longitudinal study. The gap between accuracy scores for *a* and *the* narrowed by the end of the two-year study, but the accuracy scores for *a* reached 90% only in the [+article] L1 group. The scores for *the* were at ceiling already in round 2 in both [+article] and [-article] L1 groups (after an average of 16 months of exposure to English). In regard to error types, the fact that *the* was used as the substitution form for the lower frequency article *a* in all rounds of the study provides further support for a usage-based account.

7.3.2 Articles as parts of constructions

The usage-based account predicts the accuracy with functional words or morphemes to partially depend on the context they are used in. With respect to articles in particular, not only the input frequency of individual articles, but also the frequency of items preceding or following them is expected to influence the production patterns. The analyses in this section attempt to answer the following questions: How are articles distributed with respect to frequent construction types in the input and in the children's output? Is article use dependent on the collocates of articles? Both analyses were aimed at determining whether there was evidence for construction-based storage of word combinations in children's lexicons.

The use of the indefinite article in existential 'there is a…' constructions

In order to address the prediction about the distribution of articles by construction types, two analyses were conducted. The first analysis concerned the possible formulaic status of the existential *there* construction in children's speech, illustrated in examples (1)–(3) in Section 7.1, and the second analysis concerned unequal distribution of accurate suppliance of the indefinite article between subject and object nouns, illustrated in (4).

For input estimates, the spoken COCA was searched and the number of indefinite and definite nouns immediately following 'there is/are/was/were' word strings was retrieved. The search query was "there is|was [a|an|the] [n]". Such a query does not necessarily return all occurrences of *there BE NP* construction, because it misses the cases where the noun is separated from the verb by, say, an adverb (e.g. *there is always a N*). Despite this limitation, the search results were

sufficient as an estimate of the article distribution in this construction. The distribution of definite and indefinite articles in *there* construction in the adult corpus was as follows: the indefinite article occurred in this construction 63,777 times (relative frequency 781 times per million words), and the definite article occurred in this construction 4,017 times (relative frequency 49 times per million words)[9]. This distribution is not surprising since the construction is meant for introducing new referents into discourse. The proportion of *a* in the construction is not 100% because of the cases where *there* is used with a literal, locative meaning (e.g. *There is the spoon - on the table*).

As for the children's corpus, there were 67 *there BE NP* contexts in the narratives, and in 100% of cases, children correctly supplied an indefinite article with the noun. Thus, there was a clear effect of the syntactic context, as children's overall mean percent correct use of *a* never reached 100% accuracy even at Round 5. Out of the 67 *there* constructions, 62 contained a past tense form of BE: *there was a N*, and the remaining 5 had the form *there is a N*. It is very likely that *there is a* and *there was a* were memorized expressions in children's lexicons. However, this by no means implies that all correct uses of *a* in the children's speech were part of a memorized expression. *There* contexts took up about one third of all the uses of *a* in the children's speech, and the earliest time when the uses of *there* constructions were found in narratives was Round 3.

The use of the indefinite article with subject and object nouns

In order to provide further support for the idea that the children were more accurate with morphemes that were part of frequent, and therefore more familiar constructions, the distribution of articles in indefinite contexts other than *there* + *BE* was analyzed, taking into account the syntactic position of each scorable indefinite noun. This was done in order to find out whether article suppliance was indeed quantitatively different in subject and object contexts, illustrated earlier in (6). The distribution by syntactic context is summarized in the table below.

[9] The counts for COCA were obtained on April 29, 2010. Since this corpus is periodically updated, current counts of articles in the spoken susbcorpus can differ.

Table 7.2

In the children's English, frequencies and percentages of target and non-target articles in indefinite contexts

	Correct *a*		Incorrect *the*		Total
Object	111	81%	26	19%	137
Subject	27	36%	49	64%	76
Total	138	65%	75	35%	213

It is clear that the children were more accurate with supplying the indefinite article with objects. Another way of looking at error distribution is to say that two thirds of all *the* misuse errors were found with subject nouns (49 out of 75).

Since the numbers in Table 7.2 represent the total frequencies of suppliance across 5 rounds of the study, it was also necessary to check how the accuracy was distributed in each round in order to see whether the subject-object asymmetry was persistent throughout the two years or the effect was limited to the early stage but influenced the total scores. Percentages of indefinite articles out of the total number of obligatory contexts with subject and object nouns are plotted in the figure below.

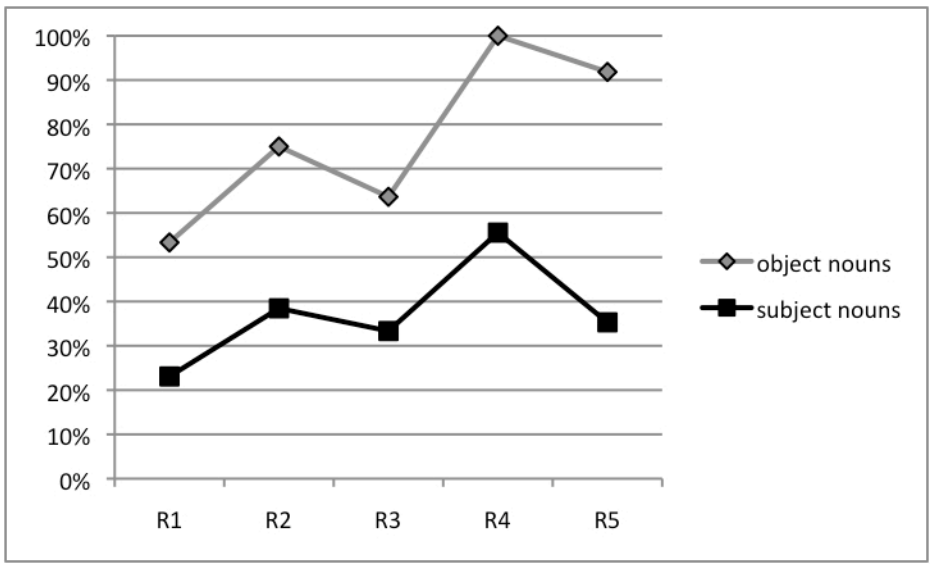

Figure 7.1. In the children's English, percent correct use of indefinite articles with subject and object nouns

There is an obvious gap between percent correct use of *a* in the two syntactic contexts that was clearly present throughout the two-year study. In fact, the gap appeared to widen towards the end of the study, as the accuracy in object contexts progressed much more rapidly.

Since the syntactic position appeared to influence the suppliance of the article substantially, it was possible that the types of verbs and prepositions that were used to introduce the object nouns influenced the accuracy of article use with these nouns. In Table 7.3, frequencies of target and non-target articles are given for the four most frequent words that preceded indefinite nouns.

Table 7.3

For indefinite nouns, children's suppliance of indefinite and definite articles in the most frequent contexts

	a(n) N	*the N*
GET	17	1
HAVE	23	1
MAKE	13	1
WITH	10	6
Other	48	17

There were very few instances of the non-target article *the* when the indefinite noun was introduced by one of these frequent words. The overuse of *the* appeared to be higher with nouns following the preposition *with*. An example of *the* overuse is given in (5a), which contrasts with correct article use (5b) in a similar context within the same story.

(5) a. There was **a zebra** with **the plane**.

b. Then a big girl elephant, she came with **a net**, a big, long net.
(SMNS, L1 Spanish, R3, 27 MOE)

In (5a), both nouns are mentioned for the first time in the story and require indefinite articles. However, the target article was supplied with the noun introduced with the words *there was*, but not with the noun introduced with the

preposition *with*. Later the same story, the indefinite article was supplied correctly with a noun following the same preposition *with*.

The distribution of articles in adult speech

In order to check whether *the* overuse was influenced by the input frequency of specific combinations of verbs or prepositions with articles, an additional analysis of adult speech was performed, focusing on the words that preceded object nouns. In COCA, search queries such as, for instance, '[get] a|an' and '[get] the' ('any form of the verb *get* immediately followed by an indefinite or definite article') were used to retrieve the number of definite and indefinite nouns in object positions. The entire spoken corpus was used for the search. The results are given in Table 7.4.

Table 7.4

In spoken COCA, frequencies of indefinite and definite articles in verb+article and preposition+article combinations[10]

	a(n) N	*the N*
GET	41,190 (504/mil)	24,789 (303/mil)
HAVE	137,987 (1,689/mil)	41,221 (505/mil)
MAKE	27,508 (337/mil)	15,494 (190/mil)
WITH	42,852 (525/mil)	80,566 (986/mil)

Given that the definite article is almost twice as frequent as the indefinite article in spoken English, the important finding is that *a* was actually more frequent than *the* when it was used with nouns following such frequent verbs as *get* and *have*. Definite nouns occurred more frequently after the preposition *with*, and this in fact was the context in which children made more the overuse errors than with nouns following verbs. Furthermore, subjects (or nouns preceding the verb, in this case) were a much more common context for the definite article: the definite article was over 10 times as frequent as the indefinite article in these contexts (2,860 tokens vs. 248 tokens). This finding again confirms that the children's choice of articles was at least partly influenced by the frequency of article forms

[10] These results were obtained on April 29, 2010.

and, importantly, also by the context of the noun with which the article was supplied.

In COCA (even in the sub-corpora of the spoken corpus), it was not possible to calculate the overall frequency of article+noun combinations before and after verbs due to the high frequency of each item in the search string. COCA search is limited to queries that return less than one million hits, and each of the slots in these combinations (article, noun, verb) would yield more than one million hits. The British National Corpus (BNC) was used instead because, due to its smaller size, it could be searched even for such frequent items as verb and noun. BNC was queried using Mark Davies' website (http://corpus.byu.edu/bnc/), which has the same interface and the same search query syntax for the American and the British corpora. In order to estimate the distribution of indefinite and definite nouns in subject positions, the corpus was searched for the frequency of nouns that followed a period "." (i.e. were at the beginning of a sentence[11]) and preceded a verb. The slot immediately following a verb or a preposition was considered as the typical object position. The search queries were as follows:

- ". a|an [nn*] [v*]"and ". the [nn*] [v*]" for indefinite and definite nouns preceding verbs;
- "[v*] a|an" and "[v*] the" for indefinite and definite nouns following verbs;
- "[pr*] a|an" and "[pr*] the" for indefinite and definite nouns following prepositions

This search does not yield an exhaustive representation of all nouns in subject and object positions because the search query misses cases when the subject noun is modified by an adjective (e.g. *A big elephant showed up*) or when there is an adverb separating the subject and the verb (e.g. *the giraffe quickly found the ball*). The search certainly has limitations, but its results of are intended to be used as a general estimate of the tendencies for article distribution. The results are given in the table below, which compares the number of definite and indefinite articles

[11] The wildcard [pu*] for "any punctuation mark" was not used because it would have included commas in the counts

with nouns immediately preceding verbs (subjects) and with nouns following verbs and prepositions (objects).

Table 7.5

In spoken BNC, frequencies of indefinite and definite articles with nouns in subject and object positions

	a(n) N	*the N*
…V	248 (25/mil)	2,860 (287/mil)
V …	103,858 (10,423/mil)	95,612 (9,596/mil)
P …	47,920 (4,809/mil)	183,564 (18,423/mil)

There is a tendency for definite nouns to occupy subject positions, in which they were over ten times as frequent as indefinite nouns. The frequencies of indefinite and definite nouns in object positions after a verb were similar, with indefinite nouns being slightly more frequent, but still this is a remarkable result, given that the definite article per se is much more frequent than the indefinite article. Prepositional phrases showed a distribution of article types that was different from the distribution in the slot after the verb, with definite nouns being much more common in prepositional phrases. This is a more general trend that corroborates the earlier result regarding the higher frequency of definite nouns after the preposition *with* (see Table 7.3).

There can be an underlying semantic reason for the uneven distribution of articles in various syntactic positions. For instance, it could be the case that indefinite nouns representing new information tend to occur more often in the object position closer to the end of an utterance, while definite nouns, or old information, are used more often at the beginning of the sentence, or in the subject position. In languages with flexible word order, the distribution of old and new information is achieved by changes in word order. In languages with fixed word order, this can be achieved via changes in the syntactic structure of the sentence.

Article suppliance in definite contexts in the children's English

So far, the discussion has focused on indefinite contexts since these were the contexts in which most of article errors occurred. As for definite contexts,

children were very accurate in supplying *the* already in round 1 of the study, when their average exposure to English was 9 months (in the [+article] L1 group, accuracy with *the* was 90%). Most of the errors in definite contexts were errors of omission in children from [-article] L1 backgrounds, whose accuracy reached 90% in round 2. As for the third type of error, namely the substitution of *a* for *the*, there were only four such article choice errors in the entire corpus of 9 children.

(6) He [= character 1] have a airplane. He [= character 2] get **a plane** from him. He threw it in the water.
(TRRK, L1 Arabic, R1, 8 MOE)

(7) Mmm the elephant throw the ball. The horse looking xxx see. He swimming. Horse get **a ball**.
(JNNH, L1 Mandarin, R1, 18 MOE)

(8) The elephant looked the airplane. And the elephant get **a airplane**.
(JNNH, R1)

These examples demonstrate the overuse of the indefinite articles with nouns that were mentioned earlier in the story and therefore required a definite article. The small number of such substitution errors and the fact that they all occurred in object contexts further supports the conclusion that children's article use was at least in part influenced by frequent input patterns, in this case, by the frequent pattern *get a(n) N*.

In sum, a more careful look at the contexts of *the* suppliance in indefinite contexts revealed that, while this error was a protracted phenomenon that was found in all children at all stages of the two-year period, *the* was not overused in all syntactic contexts equally. It is likely that the children were more accurate with the indefinite article in the constructions where the article is more frequent, parallel to the effect that was reported for *there BE N* contexts.

7.3.3 Longitudinal analyses of article use in constructions

In this section, data from individual children will be considered in detail, with a focus on the children's progress over time, with the aim of following the process of developing productivity with articles as parts of constructions, looking at the way article suppliance was influenced by their collocates and whether articles initially came in as parts of lexically specific constructions. Group data already revealed that *the* was acquired before *a*, which can be interpreted as evidence for the role of input frequency in the acquisition order. It was also established that, in indefinite contexts, children's choice of articles was influenced by the syntactic position that the NP occupied (subject or object), as the children were more accurate supplying the indefinite article in object positions. The analysis will follow up these accuracy patterns in individual children over time.

JNNH (L1 Mandarin)

First we consider data from JNNH, a Mandarin Chinese speaker, who was 6;11 at the beginning of data collection, and had had 18 months of exposure to English. JNNH had had more exposure than all other children, but at the onset of the study her use of articles was very inconsistent. Articles were often omitted and *the* was incorrectly used in indefinite contexts with subject nouns, which is illustrated in (9) and (10). Consider the use of articles with indefinite nouns *elephant* and *horse*:

(9) Mmm, **the elephant** throw **the ball**. […] **Horse** get **a ball**.

(10) The elephant looked **the airplane**. And the elephant get **a airplane**.

At this stage, *the* was used both in subject and object contexts. However, *the* was not used as the universal substitution form in all contexts, as the child sometimes incorrectly supplied the indefinite article with definite nouns. Examples (9) and (10) also demonstrate inconsistent use of articles with object nouns *ball* and *airplane*.

After 2;6 years of exposure to English (Round 3), JNNH started using *a* correctly with some nouns, but only in object positions, as illustrated in (11). However, the use of articles with objects was still inconsistent. All subject nouns were marked with the definite article even in indefinite contexts, which is the case in both (11) and (12).

(11) The swimming elephant got **a net**. (Meaning 'A lifeguard got a net')
(12) The elephant want to play with **the ball**.

By the end of the study, when JNNH had had about 3 years of exposure to English, she was using indefinite articles correctly in various syntactic positions in a sentence: in subject positions and in object positions following a verb, as in (13a), and in prepositional phrases, as in (14). However, the use of indefinite articles was still inconsistent because the child sometimes marked indefinite subject nouns with a definite article, as in (13b). The sentences in (13a-b) represent the first mentions of the two characters in the story, thus the use of the definite article in (13b) is ungrammatical.

(13) a. **A penguin** or something, he wanted to build **a sandcastle**.
b. **The rabbit** wanted to help.

(14) He found the pig pulling his wagon **with a balloon** in it.

The inconsistency with article suppliance in indefinite contexts was still found in the last round of the study (3;6 years of exposure). At this time, the child used articles correctly with object and subject nouns, which is illustrated in (15). However, she still occasionally used *the* with indefinite subject nouns, as shown in (16), which represents the beginning of the story and requires an indefinite noun to be used.

(15) Then **a woman** elephant came by. And she had **a net**.

(16) **The dog** # was building a castle.

SBST (L1 Spanish)

Next we consider article use in SBST, a Spanish-speaking child who was 5;1 years at the beginning of the study and had had 15 months of exposure to English. In the first round, the child did not produce any articles at all, as he used only pronouns to refer to most characters in the story. The only noun that the child used did not have any article supplied at all:

(17) They wanted to go swimming. Then after he go in, **elephant** went go!

Pronouns were initially used by many children in story-telling at the early stages of the study, especially by those children who had the least exposure to English. Perhaps it was a strategy to simplify the task, as some of them had difficulty remembering the right English words for the characters they saw in the pictures. SBST continued overusing pronouns for many referents in the stories later on, after 1;9 years of exposure, as illustrated in (18) and (19).

(18) **They** want [//] wanted to make **a castle**, a castle-sand. And then **they**'re make one.

(19) And then **they** said to the doctor, "Could we have **a balloon**?" And then **they** both got **a balloon**.

In the stories given in (18) and (19), the characters were never referred to with nouns at any point. Note that, at this stage, the child started using the indefinite article with object nouns, and supplied it correctly (e.g. *a castle*, *a balloon* in the examples above). When he did use nouns to refer to characters, *the* was overused with indefinite subject nouns. The subject-object asymmetry is illustrated in (20).

(20) They had **a balloon**.
And **the rabbit** came.
And he said, "I want **the balloon**".

The indefinite article is supplied correctly with the object noun *balloon*, but the non-target article *the* is used with the subject noun *rabbit* in the next utterance. After 2;3 years of exposure, the child continued using *a* accurately with object nouns and overusing *the* with subject nouns:

(21) Um, **the little zebra and elephant** was going swimming. And the elephant took **a ball** with him.

At this stage, the first use of *there* construction was found, and the nouns that were previously incorrectly marked with definite articles, shown in (21), were supplied with an indefinite article correctly this time:

(22) Um, there was **an elephant** and **a zebra**.

In the last two rounds (2;10-3;4 years of exposure), the child used the indefinite article correctly in *there was* contexts, as well as in object contexts :

(23) Once there was **a girl** and **a boy**. And um the boy wanted **a balloon**.

At this stage, the child was using *a* correctly not only as a part of *there was a…* construction, as in (24a), but also with subject nouns, as in (24c).

(24) a. First there was **a giraffe** and **a zebra**.
b. And they throw **the ball** into the water.
c. And then **a guy** came running.

However, article suppliance was still inconsistent even at this advanced stage, as the child occasionally overused *the* with indefinite nouns (24b).

MRSS (L1 Mandarin)

The Mandarin-speaking child MRSS (5;0 years old at the onset of the study), had had only four months of exposure in Round 1, the least exposure to English of all the children. MRSS didn't use any articles in Round 1, but used the word *one* instead.

(25) One time **one girl elephant** um bouncing he ball. <He # he he> [/] he saw a [//] **one # kind of horse**.

The use of *one* could be a result of transfer from L1 (the Mandarin word for *one* can be used to mark indefiniteness of nouns), but a similar pattern was not found in other Mandarin- or Cantonese-speaking children in the study. It could be due to the fact that MRSS was at a very early acquisition stage in her English, as she had had only 4 months of exposure to English at the onset of the study, but it could also be an individual characteristic of her English.

Six months later, MRSS was still using *one* to mark indefinite nouns, which is demonstrated in (26a) and (27a). However, by this time the child was also using articles in certain contexts. Similarly to other children, accurate indefinite articles first appeared only in object contexts, as in (26b) and (27a). The child appeared to mark indefinite subject nouns with either *one* or *the*, as in (26a), (26c), and (27a,b).

(26) a. One [/] **one girl donkey** is playing outside.
b. He build **a castle**.
c. And **the big bunny** see it.

(27) a. **One doggy** he [/] he buy **a cool new balloon**.
b. **The bunny** saw that.

The use of *one* with indefinite nouns disappeared by Round 3 (1;4 years of exposure). By this time, MRSS was following the pattern that was found in other children at this stage, namely using *a* correctly for indefinite objects (28, 29a), misusing *the* with indefinite subjects (30), and also using *the* correctly with definite subjects and objects (29a,b).

(28) **The giraffe** saw **a beautiful ball**. And **the elephant** was playing with it.

(29) a. And another elephant girl have **a net** to get **the airplane**.

b. And **the big elephant** can't get a airplane.

Example (29b) is an instance of the misuse of the indefinite article in a definite context (*get a airplane*), which was not very common in the children's speech. Importantly, it occurred in the context of the verb *get*, which often introduced indefinite objects in the children's speech, as well as in the adult corpus. Recall that other examples of such overuse of *a* were found in JNNH in precisely the same contexts (see examples (9) and (10) earlier in this chapter)

The final stage of article use (2-2;4 years of exposure) that was observed in the study was the stage when the child used indefinite articles both in subject and in object positions. Articles were supplied in object positions accurately, but the child was still inconsistent in the use of indefinite articles in subject positions when referring to new characters in the stories:

(30) One day **a dog** was making **a castle**. **The rabbit** saw it.

(31) **The elephant** he [//] she was bouncing **a ball**. Then **a giraffe** came and talked to her.

For the subject nouns in (30), the indefinite article is supplied correctly in the first utterance of the story, but not in the second. In the case of subject nouns in (31), *the* is overused in the first utterance (*the elephant*), but an indefinite noun is used correctly in the second utterance (*a giraffe*), indicating that even after 2;4 years of exposure, MRSS was still using articles inconsistently in some contexts, similarly to JNNH and SBST.

7.3.4 Stages of development in the acquisition of the article system

At the earliest stage (Round 1, average 10 months of exposure), only the definite article was used or articles were often omitted altogether. The children often resorted to using pronouns to refer to all characters perhaps because they had difficulty remembering the right name for the animals in the pictures. MRSS, a Mandarin-speaking child, started off using *one* as a default article. As no other child was at such an early stage in the study (MRSS had only 4 months of exposure to English), it is not possible to confirm whether this is a common initial stage for L1 Mandarin speakers.

At the next stage (Rounds 2-3, average 15-20 months of exposure), a clear change in article use was observed in all children. The definite article was supplied with all subject nouns regardless of whether they were definite or indefinite, and the indefinite article was used correctly only with object nouns, especially in sentences such as *X get a Y* and *X had a Y*, which was also a very common construction in adult speech. This finding is a clear indication that the children were influenced by input frequency, since the more frequent *the* was the substitution form, and that the emergence of the indefinite article was piecemeal.

In Round 3 or 4 (20-27 months of exposure), most children started using the existential *there BE N* construction, and the indefinite article was always supplied correctly in this construction from this stage on. Articles were still supplied inconsistently with subject nouns, further confirming that productivity was not established instantaneously, but over a long period of time. The children learned the indefinite article in a piecemeal fashion, gradually increasing the number of constructions it was used in, which would be unlikely if they had established an abstract category D and accessed the features [±definite]. Finally, at the last stage of the study (34 months of exposure), accurate article suppliance was found with object nouns and most of subject nouns. The overuse of *the* was still found with subject nouns even in the last round of data collection in most children.

7.4 Summary

The analyses in this chapter investigated the theoretical issues of the role of input frequency and the growth of productivity in child L2. The question that pertained to the influence of input frequency on the children's accuracy with articles was the following: *How does the frequency of* the *and* a *in the input compare with that in the children's speech?* The more frequent article *the* was acquired much earlier than *a*. The children were more accurate supplying the definite article and used it as the substitution form in indefinite contexts. The comparison of the distribution of articles by construction type in the input and in the children's output demonstrated that the children's article use was influenced by the patterns in adult speech. The children were 100% accurate in existential constructions where the indefinite article is a much more frequent option than the definite article in adult English. The children were also more accurate with supplying the indefinite article with nouns in object positions, especially in direct object positions immediately after a verb, which was also the position favored by indefinite nouns. Finally, with nouns in subject positions, the erroneous use of *the* was the most common error, and an analysis of adult speech demonstrated that definite nouns are placed in subject positions much more often than in other syntactic positions, again indicating that the children were influenced by the distribution of indefinite and definite nouns in the input.

The other question that this study addressed was the following: *Is there evidence for piecemeal acquisition of the article system?* The evidence was present in the way articles initially were used correctly only as parts of a lexically specific construction and in the way article use was dependent on the collocates of articles. The analysis of the children's suppliance of articles with object nouns revealed that the word that precedes the NP in question appears to influence the choice of article in that NP. For instance, the children were very accurate with supplying the target indefinite article in *get NP* phrases (e.g. *She got a net*), and even overused the combination *get a N* for nouns that required a definite article. Interestingly, the children were less accurate with supplying the indefinite article in the PP *with a N*, and the analysis of adult spoken English revealed that

indefinite nouns occur more frequently in NPs following the word *get* than the word *with*.

Importantly, it was not the case that all children's uses of articles were parts of memorized routines throughout the longitudinal study. The fact that they were 100% accurate using *a* in existential *there* constructions does indicate that *there is/was a*... was possibly a formulaic expression in the children's stories. However, while the indefinite article came in as part this construction, the variety of contexts of its use increased gradually. The conclusion that they did not entirely rely on frequent word combinations is also supported by the finding that *a* was not overused in definite contexts, with only four exceptions. Frequency effects need to be considered together with other factors. Children relied on frequent lexically specific constructions when supplying the more semantically complex indefinite article, whereas the definite article was acquired with ease. This finding can also be taken as support for the usage-based model of acquisition, since we found that members of the same theoretical category such as article demonstrated very different acquisition paths.

Taken together, the findings also shed light on the nature of the most common article choice error found in learners of L2 English, namely the overuse of *the*. The analyses clearly demonstrate that *the* was not overused in all syntactic contexts at the same rate, which is a strong piece of evidence for piecemeal acquisition. It is likely that the children started using *the* earlier than *a* and chose *the* as the substitution form due to its overall high frequency (it is in fact the most frequent word in English), but the contexts in which the children were more accurate supplying *a* were precisely the contexts in which the indefinite article was more frequent in adult speech (e.g. with direct object nouns). Thus, the preference for indefinite nouns in these contexts in the adult language 'protected' them from *the* overuse in the children's speech. This finding supports the approach to the role of frequency advocated by Bybee (2007) that not only the token frequency of individual words (in this case, *the* vs. *a*) should be taken into account, but also frequencies of longer more specific strings (e.g. *with the*… vs. *with a*…, and *get the*… vs. *get a*…).

CHAPTER 8. CONSTRUCTIONS IN CHILD L2 ENGLISH: THE AUXILIARY VERB SYSTEM

8.1 Introduction

This chapter continues the investigation of the role of input frequency in children's grammatical development over time and piecemeal acquisition of constructions in L2 acquisition. These issues have been studied intensively in L1 acquisition but less so in child L2 acquisition. Even though this approach has only been applied to L1 acquisition in earlier studies, its interpretation of the role of input factors should hold for child L2 learners as well as L2 learners in general. In the present study, L2 children's progress will be followed in order to test whether their accuracy is influenced by the frequency of auxiliary forms in the input and whether L2 children rely on frequent unanalyzed combinations in building more complex constructions, just like children learning a L1 do. Similarly to the analysis of articles in Chapter 7, the aim of the present chapter is to analyze the children's use of auxiliaries from the usage-based perspective in order to find out whether the frequency distribution of grammatical morphemes in the input is reflected in the order of acquisition, to find out whether auxiliary use is influenced by their collocates in constructions, and to follow the emergence of productivity in children's use of auxiliary constructions over time. Increasing productivity will be analyzed by following the development of the subject and the main verb slots in the earliest schemas. Evidence for piecemeal acquisition will also expected, i.e. different forms of the same auxiliary will come in as parts of separate constructions or at different times in the development.

8.1.1 Background to the study

A growing number of studies provide evidence for the role that input frequency plays in children's language development and that their productivity develops gradually (see Chapter 3 for review of the literature). With respect to auxiliary verbs in particular, numerous studies have found evidence for the effect of input distribution on children's accuracy. For instance, accuracy with auxiliary

BE in obligatory contexts was higher for the more frequent form *is* than for *are*, and it was higher when auxiliaries followed high-frequency subjects. The order in which particular subject+BE combinations are acquired reflects their frequency in the input as well. For example, children were significantly more accurate with supplying *is* in the frequent combinations *it's* and *he's* than in other combinations, such as *she's* or *that's*. With regard to the acquisition of auxiliaries in questions, previous studies reported that children produced significantly more correct *wh*-questions with *is* than with *are* in spontaneous speech. Overall, previous usage-based research on L1 acquisition suggests that the use of each auxiliary is learned in a piecemeal fashion, that there is a continuum of constructions with varying lexical specificity, and that patterns of use in children's speech are closely related to the patterns in the language they hear. The analysis presented in this chapter aims to explore the degree to which the L2 children's early speech is based on lexically specific constructions and by exploring the interaction between auxiliary use and the properties of the input. The analysis will target the growth of productivity of L2 children's auxiliary constructions over time (i.e. the growth of variability of the slots in N+BE+V-ing, DO+N+V and N+DO+*not*+V constructions) and how the input frequency of auxiliaries influences this growth. It will use elicited and spontaneous longitudinal data from 9 children with various L1 backgrounds.

8.1.2 Research questions and predictions

In order to explore the lexically based nature of children's early speech and the properties of the input that influence early auxiliary use and non-use, the following research questions will be answered:

(1) Does the order in which children acquire specific auxiliary forms reflect their relative frequencies in the language children hear? Usage-based studies found that accuracy rates vary among different auxiliaries and even among different inflected forms of the same auxiliary, depending on the frequency of the form in the input (e.g. Theakston et al. 2005; Theakston & Rowland, 2009a,b). Thus, it is predicted that (a) higher frequency forms of BE and DO will have

higher rates of suppliance (e.g. accuracy rates will be higher for *do* than for *does*, if *does* is the more frequent form in the input); (b) the range of inflected forms for each auxiliary is expected to grow gradually over time with more exposure; (c) in regard to error types, higher frequency forms will be used as substitution forms for lower frequency forms at the early stages of acquisition.

In order to investigate the issue of piecemeal acquisition and emergent productivity in auxiliary constructions, auxiliaries will be analyzed not in isolation, but together with their collocates in constructions. The following two questions will be addressed:

(2) What are the effects of the subject type? The usage-based account predicts context-specificity in the early use of auxiliaries. In other words, not only the frequency of individual auxiliaries, but also the frequency of items preceding or following them is expected to influence the production patterns as follows: (a) Fully lexically specified combinations (for example *I'm* or *he's*) have high token frequency and are likely to be learned early and stored as linguistic wholes. Thus, combinations of pronoun+auxiliary or auxiliary+pronoun are expected to be produced accurately early on; (b) Since token frequency of various NP+auxiliary combinations is lower due to variation in the subject slot, suppliance of auxiliaries in such combinations is likely to be lower than in fixed pronoun+auxiliary combinations. In order to use auxiliaries with NP subjects correctly, children are required to develop a more abstract schema. As the development of abstract schemas is assumed to take place gradually, auxiliary suppliance in schemas with variable slots will initially be lower than in fixed pronoun+auxiliary combinations.

(3) What is the process of developing abstractness in the subject position and is there evidence for piecemeal acquisition in auxiliary constructions? The analysis will look at the intermediate stages between fully memorized constructions and fully productive constructions. It will be done in two ways: (a) The growth of abstractness can be seen in the development of flexibility of the subject slot in the earliest schemas. Schemas are predicted to come in initially with a fixed subject, for instance *I* in [*I* + *don't* + V] construction. Over time,

subject variation is expected to appear in the initially lexically-specific slots. (b) Different forms of the same auxiliary may often come in as part of separate constructions, for instance *do* in negative declaratives can appear much earlier than *do* in interrogative utterances. Over time, children are expected to link separate constructions to form one, more abstract, construction.

8.2 Method

8.2.1 Participants

The children whose data were used for the present analysis were the same as those selected for the usage-based study of the article system in Chapter 7 (see Section 7.2.1, Table 7.1). Recall that there were 3 children in each L1 group: Arabic, Spanish, and Mandarin Chinese. Data were collected in freeplay situations as well as elicitation tasks every 6 months for two years. From each session, two types of data will be used in the analysis, namely the use of auxiliaries BE and DO in the Test of Early Grammatical Impairment and in spontaneous speech, described in Chapter 4 (Sections 4.2.1, 4.2.3). The test provided elicitation contexts only for third person singular and plural forms of auxiliaries. In total, the BE/DO probe comprised 11 attempts of DO-questions, 6 attempts for BE-statements, and 6 attempts for BE-questions. Spontaneous speech samples were obtained during play sessions and were manually coded for the use of several target morphemes, including auxiliaries BE and DO.

8.2.2 Data Analysis

In spontaneous and elicited speech, obligatory contexts for BE+*ing* constructions and DO insertion were coded. Each auxiliary was coded for relevant person, number, and tense features. The structural and morphological context of use was also coded, namely the type of utterance (declarative or interrogative), type of subject (nominal or pronominal), main verb, and the form of auxiliary used (e.g. *am*, *are*, or *is*; *do* and *does*).

Coding of structures with DO

Coding of structures with DO elicited in TEGI and taken from spontaneous speech samples followed the same steps as described in Chapter 4 (Section 4.2.3). Use of DO was scored as correct when the auxiliary was used with correct tense and person agreement, and the main verb was uninflected. Commission (wrong form) errors were uses of DO with wrong agreement (DO for DOES or DOES for DO), wrong tense (DID for DO), or wrong auxiliary (IS for DOES). Double marking errors refer to utterances with DOES in which the main verb was inappropriately inflected with 3sg *–s* and utterances with DID in which the main verb was inflected for past tense. While TEGI elicited only questions with third person forms of auxiliary DO, in spontaneous speech there were instances of DO with negation. In negative sentences, omission errors refer to sentences with the negation marker *no*/*not* and a main verb. Example (1) below provides illustrations of all types of errors.

(1)	a.	Is the kitty needs more milk? (YSSF, L1 Arabic, R2, 16 MOE)	commission
	b.	Does he likes yellow? (DVDC, L1 Spanish, R2, 15 MOE)	double marking
	c.	He no give it to her. (SMNS, L1 Spanish, R1, 13 MOE)	omission
	d.	Here I don't saw Pokemon. (MRSS, L1 Mandarin, R1, 10 MOE)	commission

It was not possible to precisely characterize the nature of errors in questions with an uninflected main verb, as in (2a), since they can be analyzed as omission of DO or a declarative structure with question intonation. However, it was possible to detect unambiguous omission of DO in *wh*-questions, and such instances were scored as omissions:

(2) a. You turn it [= microphone] off? (TRRK, L1 Arabic, R2, 14 MOE)
b. How you do that? (YSSF, L1 Arabic, R1, 16 MOE)
c. Hey, how this get in here? (SBST, L1 Spanish, R1, 15 MOE)
d. Why you never get it? (TRRK, L1 Arabic, R2, 14 MOE)

In the coding process, a large number of negative and interrogative utterances were identified that were very likely to be memorized in their entirety. For instance, formulaic expressions such as *I don't know* and questions such as *Do you know what?* were used in spontaneous speech very frequently by all the children. These utterances were not included in the calculation of the correct provision of auxiliaries in obligatory contexts. For instance, DNNC (L1 Mandarin) produced only two DO-question in Round 1, and they both were *Do you know?* In Round 2, 12 out of 15 questions were *Do you know what?* and the other three questions had the form [*Do you know*] + [embedded clause]. Including such questions, which were clearly memorized chunks, would not give a true representation of the child's use of auxiliary DO. The same rationale was used to exclude *I don't know* and *I don't remember.* For instance in Round 1, the children produced 144 DO-statements in spontaneous speech, and 62 of them were *I don't know* and *I don't remember*, and *I don't wanna*. For some children, these were the only instances of DO they produced in the first round. Thus, it was decided to exclude these three expressions from the analysis altogether.

Coding of structures with BE

The coding procedure followed the same steps as the one described in Chapter 4 (Section 4.2.3). Unlike in TEGI contexts that targeted third person auxiliary forms, in spontaneous speech there was more variation in auxiliary form, Examples of all types of scorable accurate responses and errors are provided in (3) for statements and in (4) for questions.

(3) BE in declaratives

correct: So then it's like <he was> [//] sponge+bob **was** like drying up. (DNNC, L1 Mandarin, R5, 14 MOE)

omission: I just eating. (TRRK, L1 Arabic, R3, 19 MOE)

commission: My mom **had** telling me . (TRRK, R5, 37 MOE)

(4) BE in questions

correct: What'**s** she staring at? (YSSF, L1 Arabic, R4, 34 MOE)

omission: What she doing girl? (TRRK, R1, 8 MOE)

commission: What **does** it doing? (YSSF, R1, 9 MOE)

double marking: What'**s** you'**re** talking about ? (TRRK, R3, 19 MOE)

The approach explored in this chapter makes different predictions for accuracy with auxiliaries depending on the particular inflected form they are used in and on the syntactic context of each instance of auxiliary. Thus, in addition to scoring children's general accuracy with auxiliaries, the scoring procedure also recorded the particular form of BE and DO that was used (e.g. *is* or *are*; *does* or *did*), the type of subject (nominal or pronominal), and the main verb.

8.3 Results

8.3.1 Distribution of auxiliary forms in adult speech

In order to explore the role of input in children's use of auxiliaries, it was first necessary to establish which auxiliary forms were more frequent in the input to children. Since it was not possible to obtain samples of spoken English input that these children receive in their homes or at school, corpora of adult spoken English were considered as models of the speech that the children hear. The frequencies of various inflected forms of BE and DO were calculated in the spoken subcorpus of Mark Davies' Corpus of Contemporary American English (COCA: Davies, 2008; see Chapter 4.2.4). The frequency data for various forms of auxiliary BE in spoken American English are summarized in Table 8.1. In this table, the frequency of each auxiliary is broken down into frequencies of specific

inflected form, in order to compare the distribution of various person and number forms (*am*, *are*, and *is*). The frequencies of third person forms are also broken down by subject type, in order to compare the frequency of noun+auxiliary and pronoun+auxiliary collocations. Past tense forms *was* and *were* will not be considered because they were very infrequent in the children's spontaneous speech.

Table 8.1

Frequencies of BE forms in spoken American English

	per million words
N + is	3,745
N + 's	1,508
pronoun + is	1,695
pronoun + 's	5,963
Total *is*	**12,234**
N + are	1,931
they + are	420
they +'re	1,116
Total *3pl are*	**3,467**
(I) am	327
(I) 'm	1,999
Total *am*	**2,326**
(you) are	246
(you) 're	1,281
Total *(you) are*	**1,527**
(we) are	407
(we) 're	1,372
Total *(we) are*	**1,779**

These particular forms of BE were chosen for comparison due to the nature of the data available. In the case of elicited speech, the only two target forms were third person forms *is* and *are*, and in the case of spontaneous speech, the most frequent BE forms that occurred were *am* and second person *are*. Since other forms of BE in spontaneous speech had very low frequencies in individual samples, it was decided to limit the comparisons of accuracy to two pairs of forms, *is* and *are* for TEGI data, and *am* and *are* for spontaneous speech data.

Part of speech tagging in COCA does not distinguish copula and auxiliary BE, and thus the frequencies for *am*, *are*, and *is* represent the tokens of both

copula and auxiliary pooled together. In the calculation of the frequencies of the same forms in the input to children learning L1 English, Theakston & Rowland (2009: 1458) also analyzed the copula and the auxiliary frequencies together, stating that input of copula BE might contribute to children's knowledge of auxiliary BE, and vice versa.

8.3.2 Children's accuracy in TEGI and spontaneous speech

Recall that in addressing the research question *Does the order in which children acquire specific auxiliary forms reflect their relative frequencies in the input?* The following predictions were formulated: On the basis of the input frequency data, higher-frequency forms will be acquired earlier than lower-frequency forms. In other words, one might expect that in third person contexts, the high-frequency auxiliary form *is* will be less error prone than the lower frequency form *are*. In first and second person contexts, auxiliaries *am* and *are* will be supplied with similar accuracy. The data from TEGI and freeplay will be analyzed separately. First, proportions of correct suppliance across the 5 rounds of the longitudinal study will be presented, due to low individual frequencies in each round. These analyses will be later followed by a round-by-round analysis of individual trajectories of development in children (in Section 8.3.6).

Table 8.2

In TEGI, overall proportions of correct provision of the plural and singular forms of BE

	are	**is**
DNNC (Ch)	0.88 (22/25)	0.91 (32/35)
JNNH (Ch)	0.79 (23/29)	0.89 (24/27)
MRSS (Ch)	0.68 (17/25)	0.90 (27/30)
DVDC (Sp)	0.80 (20/25)	0.88 (23/26)
SBST (Sp)	0.94 (30/32)	0.93 (26/28)
SMNS (Sp)	0.95 (19/20)	0.94 (17/18)
LLKC (Ar)	0.69 (11/16)	0.79 (15/19)
TRRK (Ar)	0.67 (20/30)	0.84 (21/25)
YSSF (Ar)	0.71 (20/28)	0.94 (29/31)
Mean (st.dev)	**0.79 (0.11)**	**0.89 (0.05)**

As predicted, the children were more accurate with the form of the auxiliary that is more frequent, namely with the form *is*. The data in Table 8.4 are collapsed over all rounds of the longitudinal study and represent overall accuracy with auxiliaries. It was not possible to calculate individual correct scores for each auxiliary form in each round due to low frequencies, but to get a better picture of the development over time, total accuracy scores across 9 children were calculated for each round. The results are given in Figure 8.1.

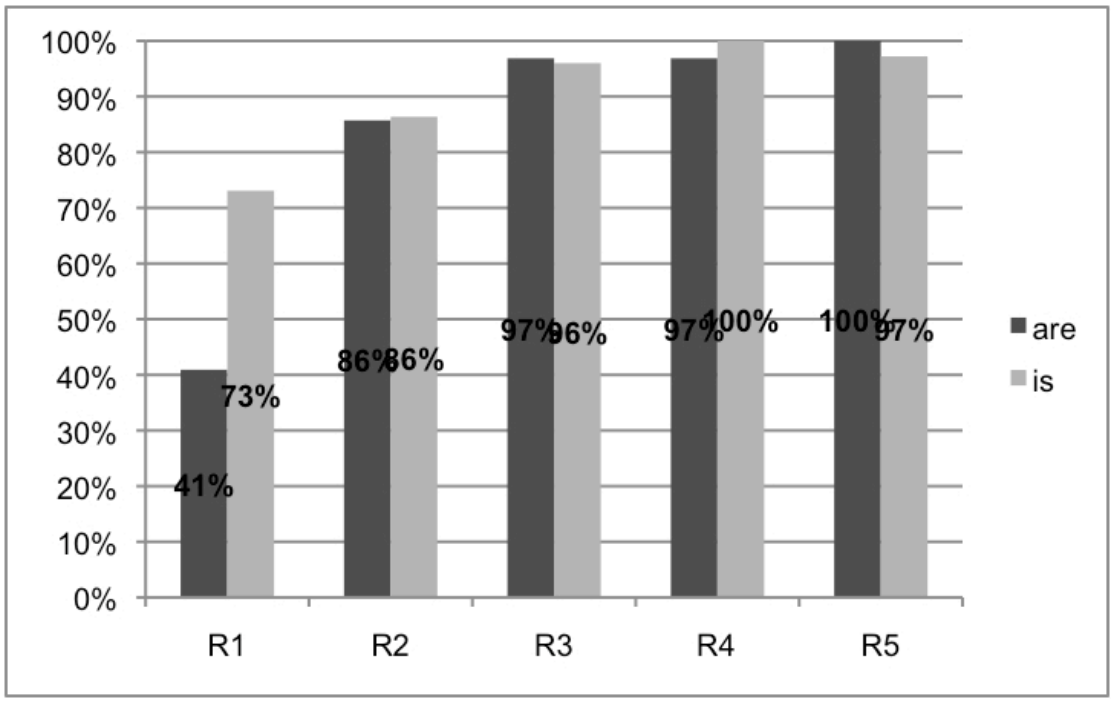

Figure 8.1. In TEGI, percent correct suppliance of 3rd person BE forms in each round

The overall difference between the accuracy with *is* and *are* clearly stems only from the children's performance in Round 1. In other words, the form *is* was acquired earlier, but the children's performance with the form *are* caught up already by Round 2, i.e. in 6 months. Comparing examples from each round, it can be observed that the plural form *are* was absent in most children's responses in the first round. At the same time, the auxiliary *is* was supplied early on. For

instance, the child DVDC used the form *is* in the first round, but the auxiliary *are* appeared only in the second round:

(5) He is crying? DVDC (L1 Spanish, R1, 8 MOE)
The moon guys crying? DVDC (R1)
Maybe they're fooling around. DVDC (R2, 15 MOE)

Another reason why accuracy with *are* was lower than with *is* could be that children often used *are* only in combination with *they*, and omitted it when the subject was a noun.

(6) They are having fun.
The bears crying? LLKC (L1 Arabic, R1, 11 MOE)

In fact, in LLKC's elicited data, the form *are* was used correctly only with the pronominal subject in all rounds of the study.

In the children's spontaneous speech, the majority of inflected BE forms were auxiliaries *am* and *are*, and since there were very few questions with BE, these forms occurred in contracted combinations *I'm* and *you're*. As stated in Table 8.1, the frequencies of these forms were similar in adult speech, with *I'm* only slightly more frequent than *you're*. Therefore, differences in the accuracy of suppliance of these forms were not expected.

Table 8.3

In spontaneous speech, proportions of correct provision of 1st and 2nd person forms of BE

	am	**are**
DNNC (Ch)	0.53 (8/15)	1.00 (9/9)
JNNH (Ch)	0.89 (8/9)	1.00 (5/6)
MRSS (Ch)	0.80 (8/10)	1.00 (6/6)
DVDC (Sp)	0.65 (26/40)	0.63 (10/16)
SBST (Sp)	1.00 (39/39)	1.00 (6/6)
SMNS (Sp)	1.00 (30/30)	0.90 (9/10)
LLKC (Ar)	1.00 (16/16)	1.00 (4/4)
TRRK (Ar)	1.00 (16/16)	1.00 (4/5)
YSSF (Ar)	1.00 (12/12)	0.87 (13/15)
Mean (st.dev)	**0.87 (0.18)**	**0.89 (0.13)**

In the 9 children taken as a group, there was no difference in accuracy with *am* and *are*. In fact, 6 out of 9 children were slightly more accurate with *am* than with *are*, which can be due to the higher frequency of *am* in the input. As mentioned earlier, the number of utterances with auxiliaries was not sufficient to calculate reliable percent correct scores for each child in each round. Similarly to the previous analysis of *is* and *are*, total accuracy scores across 9 children were calculated for each round. The results are given in Figure 8.2.

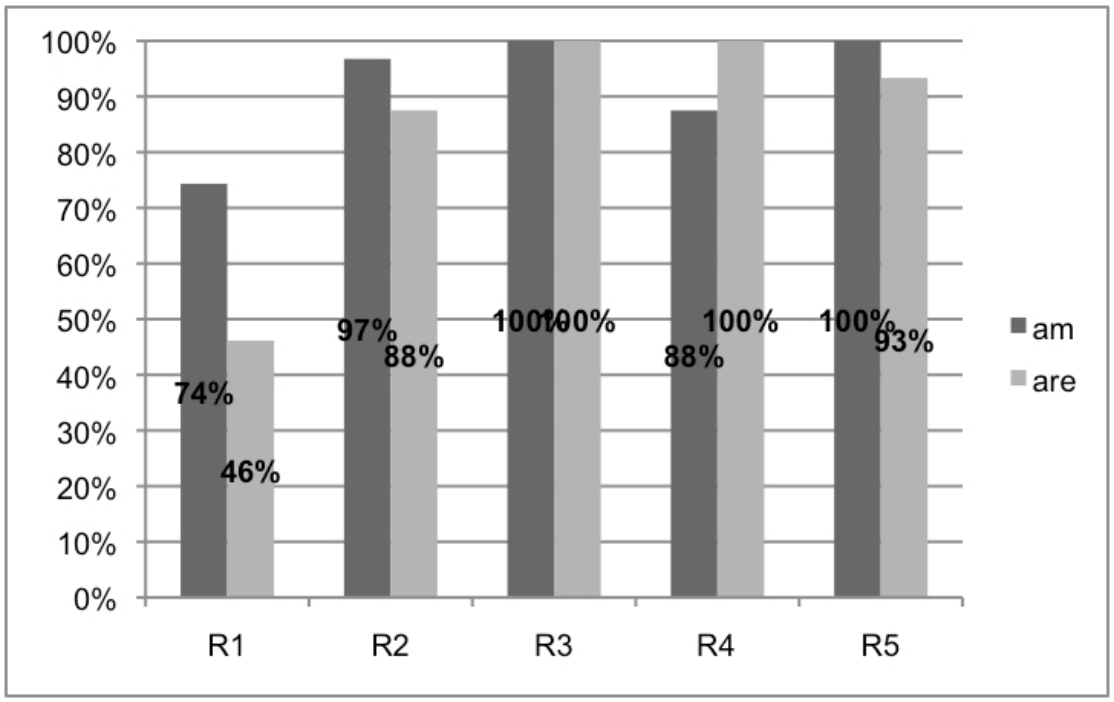

Figure 8.2. In spontaneous speech, percent correct suppliance of 1st and 2nd person BE forms in each round

The result is somewhat similar to that for auxiliaries *is* and *are*, with the substantial difference between *am* and *are* present only in the first round. The contrast between the input frequencies of *am* and *are* (2,000 vs. 1,700 per million) was not as clear as the contrast between *is* and *are* (12,000 vs. 3,000 per million), but it still appeared to influence the children's accuracy. It could be the case that factors other than frequency contributed to the facilitation of the acquisition of the form *am*. One such factor could be consistency (one function – one form

mapping), which can work together with input frequency in facilitating the acquisition of constructions (Lieven & Tomasello, 2008): the form *am* is restricted to only one subject form, namely *I*, whereas the form *are* can occur with several pronominal subjects (*we*, *you*, *they*) and also with a variety of plural nouns.

8.3.3 Analysis of errors: Auxiliary BE

This section addresses the prediction that third person contexts, the high-frequency form *is* will be substituted for the lower-frequency form *are*. In BE contexts in spontaneous speech, there were only 5 instances of substitution errors in total, that is across all children over the two years of the longitudinal study. Therefore, in the analyses in this section, only data from TEGI will be analyzed. In total, there were 31 substitution errors in BE contexts in TEGI, and 27 of these errors were found in contexts for the plural auxiliary *are*. The most common error, as predicted, was the auxiliary form *is* substituted for *are* (in 19 out of 27 contexts for *are*), illustrated in (7a) and (7b). There were also 4 errors of double marking when the auxiliary *is* was used at the beginning of a question with auxiliary *are* following the subject, as illustrated below in (8a) and (8b). The rest of the substitution errors were of the type illustrated in (9), and they were too few to show any significant trend.

(7) a. The moon guys is kissing the bug too. (DVDC, L1 Spanish, R1, 8 MOE)
b. Bears is snoring too. (MRSS, L1 Mandarin, R2, 10 MOE)
c. The moon guys was sleeping. (JNNH, L1 Mandarin, R2, 24 MOE)

(8) a. Is the moon guys are crying? (JNNH, R2, 24 MOE)
b. Is the bears are feeling better? (YSSF, L1 Arabic, R2, 16 MOE)

(9) a. Did the moon guys are laughing? (DVDC, R2, 15 MOE)

b. Does the bear feeling better? (LLKC, L1 Arabic, R3, 23 MOE)

c. The moon guys was sleeping. (JNNH, R2, 24 MOE)

8.3.4 Analysis of errors: Auxiliary DO

Substitution patterns in questions with DO

Table 8.4 presents the frequencies of DO forms from the spoken subcorpus of COCA (Davies, 2008), for negative statements and yes/no questions separately.

Table 8.4

Frequencies of DO forms in spoken American English

	per million words
don't	2,390
doesn't	477
didn't	853
Do ...?	588
Does ...?	139
Did ...?	225

Recall that the general prediction for auxiliary substitution analyses was formulated as follows. *Prediction 2:* The high-frequency forms will be substituted for lower frequency forms. On the basis of frequency distribution in adult speech, the more frequent form *do* was predicted to be substituted for the lower frequency form *does* in the present tense contexts in TEGI, and for both *does* and *did* in spontaneous speech. The pattern of children's substitution errors changed over time, so the types of substitution errors will be discussed for each round separately.

We can see that the auxiliary *is* was often used as a substitution form both in singular and plural questions, and this pattern persisted in three rounds, when the children's exposure ranged from 10 to 22.

(10) a. **Is** the bear like apples? (LLKC, L1 Arabic, R1, 11 MOE)

b. **Is** the kitty need a Kleenex? (LLKC, R1)

c. **Is** the moon guys want more juice? (JNNH, L1 Mandarin, R2, 24 MOE)

d. **Is** they like apples? (JNNH, R2)

While the prediction regarding substitution errors was only concerned with the forms of auxiliary DO, the finding that the form *is* was used as a substitution form can also be explained by the frequency account, since this auxiliary is more frequent than *do* and *does* (the form *is* occurred over 12,000 times per million words). It further confirms the finding in the previous section that the form *is* functions as a universal question marker at the early stages of acquisition. In fact, in Round 1, children demonstrated complete unawareness of the auxiliaries *do* and *does* (in interrogative sentences) and when they did not use *is* as the question marker, they resorted to asking questions by changing the intonation of the declarative sentence. In Rounds 2 and 3 (16 – 22 months of exposure), auxiliary *do* appeared as the more frequent substitution form in *does* contexts, as expected.

(11) a. **Do** the kitty likes yellow? (SBST, L1 Spanish, R2, 21 MOE)

b. **Do** the bear like milk? (DNNC, L1 Mandarin, R2, 14 MOE)

Interestingly, from Round 3 on, for some children *does* becomes the substitution form in questions with a plural subject, which goes against the prediction made on the basis of the frequency of *does* in the input. One possible explanation could be that after children begin to use the form *does*, they overuse it as the marker of plurality, perhaps attempting to make the auxiliary 'agree' with the plural subject.

(12) a. **Does** the moon guys like orange juice?

b. **Does** the moon guys wants more juice?
(DNNC, L1 Mandarin, R3, 20 MOE)

In sum, the input frequency predictions were confirmed only for the early stages of acquisition (Round 1 and Round 2, 10–16 months of exposure), when more

frequent auxiliary forms *is* and *do* were substituted for the less frequent form *does*.

Substitution patterns in negative sentences with DO

The substitution of the auxiliary *is* for *do* and *does* at the beginning of questions was found only in elicited speech. The substitution pattern was different in declarative structures with DO in spontaneous speech. While in questions *is* was the default question marker at the early stages of acquisition, in declaratives with negation, *don't* was the default substitution form. Out of 20 substitution errors found in declaratives with DO, 15 were examples of *do* substituted for *did*, as in (13), or *do* substituted for *does*, as in (14):

(13) *do* for *did*:

a. Here I **don't** saw Pokemon. (MRSS, L1 Mandarin, R1, 4 MOE)

b. I was learning fast <then I was> [//] then I **don't** need those wheels.
(SBST, L1 Spanish, R1, 15 MOE)

c. Of my pretend birthday I **don't** got it yet. (SBST, R2, 21 MOE)

(14) *do* for *does*:

a. Reem **don't** come play with me (TRRK, L1 Arabic, R1, 8 MOE).

b. Yeah <and and> [/] and &do she **don't** beat me. (TRRK, R1).

c. Because <he don't> [/] he **don't** likes. (SMNS, L1 Spanish, R2, 13 MOE)

Thus, the general prediction made at the beginning of this section was confirmed: when the children failed to supply the target auxiliary forms, they were more likely to use a more frequent form as a substitute. However, this frequency effect was observed only at the early stages. The substitution pattern for the forms of auxiliary DO appeared to reverse in some children, with the less frequent form *does* being substituted for *do*, thus overriding the frequency pattern.

8.3.5 Analysis of accuracy by subject type

In the analyses in this section, the following research question is addressed: *Is there an effect of the subject type on children's use of auxiliary verbs?* The following predictions were tested: (1) Fully lexically specified combinations (for example *I'm*) have high token frequency and are likely to be learned early and stored as linguistic wholes. Thus, combinations of pronoun+auxiliary or auxiliary+pronoun are expected to be produced accurately early on. (2) Since token frequency of various NP+auxiliary combinations is lower due to variation in the subject slot, suppliance of auxiliaries in such combinations is likely to be lower than in fixed pronoun+auxiliary combinations. In order to use auxiliaries with NP subjects correctly, children are required to develop a more abstract schema. As the development of abstract schemas is assumed to take place gradually, auxiliary suppliance in schemas with variable slots will initially be lower than in fixed subject+auxiliary combinations.

Spontaneous speech data were used for the analysis. For each child, percent correct scores for auxiliary BE were calculated separately for utterances with a nominal and a pronominal subject. Individual scores and means are provided in Table 8.5 below.

Table 8.5

Proportions of correct suppliance of auxiliary BE forms for two subject types (across all 5 rounds)

	NP	**Pronoun**
DNNC (Ch)	0.80 (37/46)	0.88 (45/51)
JNNH (Ch)	0.83 (38/46)	0.86 (31/36)
MRSS (Ch)	0.76 (29/38)	0.85 (39/46)
DVDC (Sp)	0.79 (30/38)	0.70 (64/91)
SBST (Sp)	0.93 (41/44)	0.96 (101/105)
SMNS (Sp)	0.93 (28/30)	0.93 (106/114)
LLKC (Ar)	0.40 (10/25)	0.88 (53/60)
TRRK (Ar)	0.59 (20/34)	0.87 (67/77)
YSSF (Ar)	0.80 (36/45)	0.86 (75/87)
Mean (st.dev)	**0.76 (0.17)**	**0.87 (0.07)**

In order to see how persistent the contrast in accuracy was over time, total accuracy scores across 9 children were calculated for each round of the longitudinal study, similarly to the analyses in Section 3.2. The results are given in Figure 8.3.

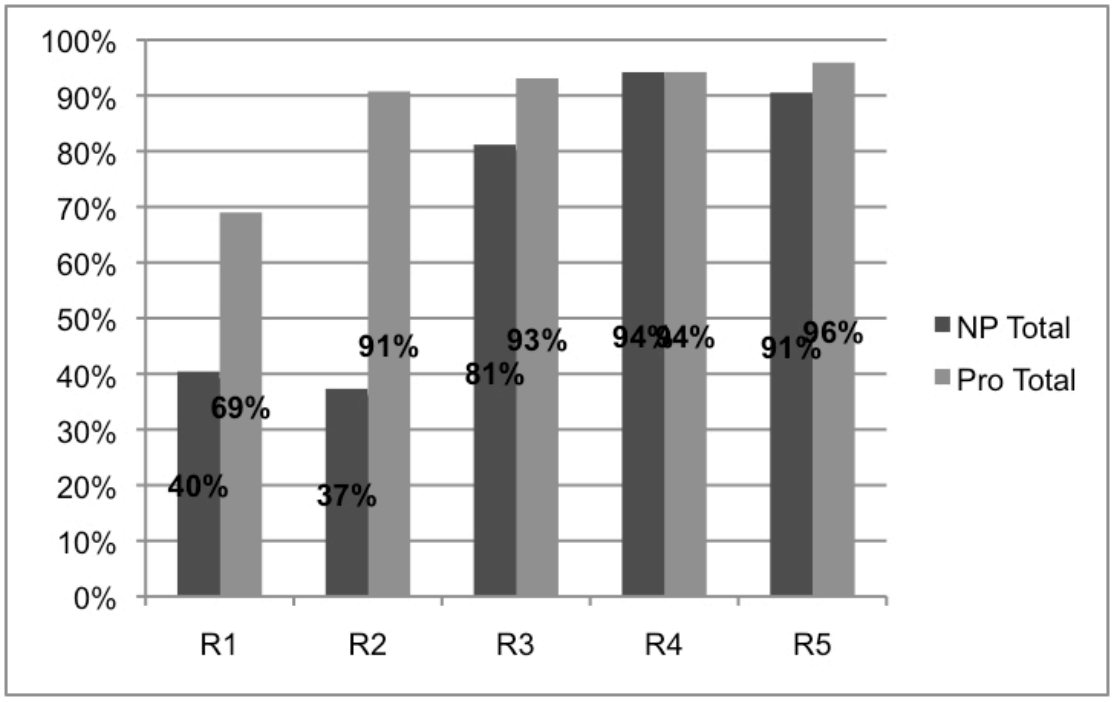

Figure 8.3. In TEGI, Percent correct suppliance of auxiliary BE with nominal and pronominal subjects in each round

As expected, children were more accurate with providing auxiliary BE in combination with a pronominal subject, and the contrast in accuracy was more lasting than the contrast based on the auxiliary form. The children were more accurate with auxiliaries followed by pronominal subjects in the first two rounds, and the contrast diminished in Round 3 (in approximately a year from Round 1).

Consider the following examples of auxiliary suppliance with nominal and pronominal subjects. In the contexts with nominal subjects, auxiliaries were often omitted:

(15) a. Then the police gonna take them. […] And, and they'**re** gonna take them. (LLKC, L1 Arabic, R1, 11 MOE)

b. Nobody talking to you. […] Zaynet telling me. […] I'**m** telling. (LLKC, R2, 17 MOE)

c. And then like one of girls gonna get candy. […] And then I'**m** gonna be Yugi. (SBST, L1 Spanish, R2, 21 MOE)

d. The camera looking only me? […] Puppy, what **are** you doing over there? (DVDC, L1 Spanish, R1, 15 MOE)

In declaratives with pronoun+auxiliary combinations, the auxiliary was cliticized onto the pronominal subject. It was the case in the examples given in (15) and is further illustrated in (16). Children's accuracy with cliticized auxiliaries could be an indication that subject+auxiliary combinations were used as memorized wholes.

(16) a. She's holding a apple. (JNNH, L1 Mandarin, R1, 24 MOE)

b. But we're not going to the fireworks. (JNNH, R5, 42 MOE)

c. She's learning lot of English now. (MRSS, L1 Mandarin, R1, 4 MOE)

d. But I'm gonna have a little brother. But I'm not gonna change diapers. (SBST, L1 Spanish, R2, 21 MOE)

e. He's [//] she's just like a same thing what you're talking. (TRRK, L1 Arabic, R3, 19 MOE)

Interestingly, the pattern of higher accuracy in pronominal subject contexts was not present in all 9 children. Table 8.5 demonstrated that three children were overall more accurate with auxiliary BE in NP subject contexts than in pronominal subject contexts, and these three children were all from L1 Spanish background. Recall from the discussion of children's L1 backgrounds that Spanish has an auxiliary *estar* 'be' that can form part of a progressive construction that is very similar to the English one. It is possible that due to the

facilitating effect of their L1, Spanish-speaking children achieved productivity with the English auxiliary BE at the earliest stage of L2 acquisition, and thus provided it in obligatory contexts regardless of the subject type (recall that, in Chapter 6, a study of larger groups of children found that the Spanish-speaking children achieved higher accuracy scores in both declaratives and questions with BE, compared to Chinese-speakers whose L1 doesn't have similar structures).

8.3.6 Development of constructions: declaratives and questions with BE

The analyses in Sections 8.3.6–8.3.9 will address the question: *What is the process of developing schematicity and generalization in lexically specific auxiliary constructions?* (a) Increasing productivity can be seen in the development of the subject and the main verb slots in the earliest schemas. Over time, variation is expected to appear in the initially item-specific schemas. (b) Different forms of the same auxiliary may come in as separate frames or at different times in the development (piecemeal acquisition). In this section, data from four children will be analyzed: Arabic-speaking children TRRK and YSSF, and a Mandarin-speaking child DNNC, and a Spanish-speaking child DVDC. These four children were chosen because they produced a substantial number of utterances with auxiliary BE contexts. Both spontaneous and elicited data will be analyzed. A case-by-case analysis will not be done for all the 9 children, as only the children who have a high number of contexts were included because their data are likely to yield analyzable patterns.

Data from YSSF

In Round 1, the verb *is* was used both with pronouns and nouns, and it was used as the substitution for *are*:

(17) a. She's sleeping.
b. The bears is sleeping.

The plural form *are* was used correctly, but only with the pronominal subject *they*, suggesting unproductive use:

(18) They're sleeping up and down. They're jumping.

There was also a contrast between the use of positive and negative auxiliary forms. While the child did use both *is* and *are* correctly, as in the examples above, he did not produce the same auxiliaries correctly in negative forms, indicating that the constructions with cliticized negative forms such as *isn't* and *aren't* were acquired later than the regular constructions.

(19) The bears no [= not having fun]. The kitty is no [= not jumping].

At the same stage, the child did not produce any correct questions with BE, indicating that an abstract construction with subject-auxiliary inversion was not yet present.

(20) a. The bears sleeping?
b. The bears crying?
c. The kitty's sleeping?

Examples of *wh*-questions in the spontaneous speech sample also indicate that, while inversion was occasionally produced correctly, the fully abstract construction was not established because of inconsistency in auxiliary form choice. Within the same session, correct use (21a) was followed by omission (21b), in turn followed by commission error (21c):

(21) a. Why **is** it not taking a picture?
b. Why it not taking a picture then?
c. What **does** it doing?

In Round 2, the form *are* was used with both pronominal and nominal subjects, suggesting that, at this stage, the child was using a construction with variable slots both for the subject and the auxiliary verb and was aware of the choice of auxiliary depending on the number of the subject.

(22) a. They're snoring very loud. The bears are still doing it.
b. The kitty is snoring louder.

We can also observe development of the question construction, as there were several examples of correct questions with auxiliary *is*, but no correct questions with *are*:

(23) a. Is the bear resting?
b. Is the bears are feeling better?

It could have been the case that the child had two constructions used for question formation: [is + N + V-ing] and [is + [declarative sentence]], which resulted in errors. In spontaneous speech, a bigger variety of BE forms was observed than in the first round. At this stage, the auxiliary was used correctly in the present and past tense with a variety of subjects. However, suppliance was not always accurate in past tense contexts, confirming the earlier observation that *is*, as the most frequent form, is acquired first.

(24) a. And then somebody's coming over.
b. Then # they were fighting.
c. He thought his web was working.
d. Yeah I was hide.
e. EXP: There was only one witch?
CHI: Who look around for me. [= who was looking around for me]

In Round 3, auxiliaries *is* and *are* were used correctly in declarative structures with a variety of subjects. In questions, however, there were still errors due to the substitution of *is* in plural contexts, again confirming the robustness of *is* as the universal substitution form.

(25) a. The bears are having fun.
b. They're jumping.
c. The kitty is jumping.
d. Is the bears feeling better?
e. Are the bears sleeping?

The substitution of *is* for *are* in yes/no questions can be found even in the last sample in the study, when auxiliary use was otherwise accurate:

(26) a. She's staring. What's she staring at?
b. We're watching tv like this. Are we gonna play a game?
c. Are they resting?
d. Is the moon guys crying?

In sum, the trajectory of development of auxiliary BE constructions in YSSF can be summarized as follows: the earliest form to be acquired was the third person form *is*, whereas the plural form *are* appeared to come in as part of [*they're* + V-ing] construction. Variability in the BE slot in the construction was seen from Round 2 on, when forms *was* and *were* begin to be supplied correctly. The question construction was acquired later than the declarative construction, and initially the child resorted to forming questions by changing the intonation of a declarative structure or using *is* as a universal question marker added to a declarative sentence. Productivity in the question construction, expressed in the variability of the BE form in [BE + N + V-ing] structures, came in sometime in Round 3, that is about a year later than in declarative constructions. However, the

frequency effect of the form *is* proved to be so strong that it was occasionally used as the substitution form in plural contexts even in the last data sample.

Data from DNNC

In Round 1, only one form of auxiliary BE is supplied, namely *is*, and even with this frequent form, there is a lot of inconsistency in suppliance in target contexts:

(27) a. I think bears crying.
b. Bear have a nap.
c. He's jumping.
d. Kitty is jumping.

The auxiliary *is* is used both with nouns and pronouns in contracted form, so it is not the case that it is used only as part of a memorized chunk *he's*. The child seems to have established the semi-productive construction [N + (i)s + V-ing]. However, there are many errors, in third person as well as first person contexts:

(28) a. Caterpillar eat he hat!
b. Uh # I eat some ice-cream.
c. I gonna show you something.

Such errors could be the result of the construction [N + (i)s + V-ing] competing with the [N + V] construction that is erroneously overgeneralized over various subjects and non-target contexts.

In Round 2, the form *are* starts to be used along with the form *is*. However, while is appears in a variety of contexts, *are* appears only in declarative structures and only with the pronoun *they*, never with plural nouns:

(29) a. The bear is sleeping. He is jumping, too.

b. They are having fun too.

c. The kitty is having a cold too.

d. They are having a cold.

Just like in YSSF's data, the more frequent form *is* serves as the question marker in both singular and plural contexts:

(30) a. Is the kitty laughing?

b. Is the bear are crying?

At the same time, in declarative structures, a variety of BE forms appears, in the present and past tense:

(31) a. And sometimes who's riding a bike wasn't going there so...

b. I'm gonna move to Victoria school.

Again, the data confirm that the question construction came in later than the declarative construction, and it took longer to achieve productivity in the auxiliary slot in the question construction, as initially all children seemed to associate the question function with the universal *is* marker that got attached to the beginning of a declarative utterance.

In Round 3, *are* began to appear in questions, however, sometimes it is erroneously attached to a full declarative structure, resulting in double marking:

(32) a. Are the moon guys are resting?

b. Are the moon guys crying?

In declarative constructions, *are* was used with a variety of plural subjects, suggesting that the child created an abstract construction with a variable slot for the auxiliary BE, that is filled with appropriate inflected forms:

(33) a. They are taking nap too. They're snoring.

b. But I know that we're having art again.

c. When you're going super-duper fast […] like you have to go &woo like that.

From Round 4 onwards, auxiliaries were supplied accurately in both declarative and interrogative constructions, with a variety of subjects and a variety of tenses:

(34) a. Are the moon guys feeling better?

b. When we were playing there was nobody in the house .

c. It's like Sandy was going to get the cookie.

d. EXP: What would you do if you were a police officer?

CHI: Like, like chase like people that were running from uh the law.

The trajectory of development of auxiliary BE constructions in DNNC was very similar to YSSF's: the earliest form to be acquired was the third person form *is*, whereas the plural form *are* appears to come in as part of *they're + V-ing* construction. Variability in the BE slot in the construction was seen from Round 3 on, that is later than in YSSF's data. This could be due to the difference in the amount of exposure and also the fact that at the early stage, DNNC appeared to overuse the construction [N + V] in progressive contexts, with BE, *-ing*, or both missing. In both children, the question construction was acquired later than the declarative construction, and initially the children formed questions by changing the intonation of a declarative structure or using *is* as a universal question marker added to a declarative sentence. Productivity in the question construction, expressed in the variability of the BE form in [BE + N + V-ing] structures, came in after Round 3.

Thus, the order of acquisition of BE forms appeared to follow the predictions based on frequency, since the form *is* was acquired first and remained as a substitution form over many months of development. The form *are*, which is

a lower-frequency form, came in later and appeared to take longer than *is* to generalize over various subject types.

8.3.7 Development of constructions: yes/no questions with DO

In parallel with Section 8.3.6 on auxiliary BE, this section will address the same research question with regard to auxiliary DO. *What is the process of developing schematicity and generalization in lexically specific auxiliary constructions?* (a) Increasing productivity can be seen in the development of the subject and the main verb slots in the earliest schemas. (b) Piecemeal acquisition is expected, i.e. different forms of the same auxiliary will come in as parts of separate constructions or at different times in the development. In this section, data from 3 children will be analyzed: a Mandarin-speaking child DNNC, an Arabic-speaking child TRRK, and a Spanish-speaking child SBST, as these were the children who had the most contexts for auxiliary DO in their speech.

Data from DNNC

In Round 1, *do*-support is not used. Questions are marked by changing the intonation of a declarative sentence or the form *is* used as a default question marker. In the elicitation task, children often attempted to repeat the experimenter's prompt, showing that they simply did not know how to form a question, as illustrated in the following examples:

(35) EXP: I wonder if the bears want some milk. Ask the puppet.
CHI: I wonder want some more milk? (Target: *Do the bears want some milk?*)

(36) Maybe they like apples? (Target: *Do they like apples?*)

In Round 2, plural *do* form is acquired and is used as the universal substitution form in all *does* contexts:

(37) a. Do the bears want some more juice?
b. Do the kitty likes yellow?
c. Do they need a tissue?

Thus, the child appeared to acquire construction [*Do* + N + V?] as a partially specified construction (with the fixed form *do* being the specified part). This construction showed a certain degree of abstractness because it was used with a variety of verbs, with nominal subjects and with the pronominal subject *they*. However, it is clear that the child did not realize at this stage that the form of auxiliary DO should vary depending on the number of the noun that follows it.

In Round 3 and 4, the singular form *does* was used correctly, but it also becomes the substitution form in *do* contexts:

(38) a. **Does** the moon guy**s** like orange juice?
b. **Does** the bug like corn?
c. **Does** the bug want**s** more milk?

At this stage, *does* is often used correctly with a variety of subjects and verbs, indicating productivity of the [*does* + N + V] construction. The reason for the overuse of *does* in plural contexts could be that at this stage, in the child 'realized' that there is a connection between the morphological form of the subject noun and the form of the auxiliary, and the child misanalyzed *does* as the plural auxiliary form.

In the last round of data collection, the child's questions demonstrate correct use of both *do* and *does*, with various types of subjects:

(39) a. **Does** the moon guy need a tissue?
b. **Does** he like hamburgers?
c. **Do** the moon guys like milk?
d. **Do** they like apples?

Data from TRRK

The same or similar stages of development observed in DNNC's data were observed in TRRK, whose L1 is Arabic. In Round 1, questions were either marked with the rising intonation, or with *is* added to a declarative sentence, regardless of the form of the subject:

(40) a. Kitty wants corn?
b. **Is** the kitty needs a tissue?
c. **Is** they like juice?
d. **Is** the bears want some more juice?

In Round 4, auxiliaries *do* and *does* appeared in questions, and *does* was sometimes supplied correctly with pronominal and nominal subjects (41a-b). However, *do* and *does* were also used as a substitution forms in non-target contexts (41c-d), incidating inconsistency. In fact, *do* was used correctly only in one context, with the pronoun *they*, indicating that the dependency between the number of subject and auxiliary form has not been established in [*do/does* + N + V?] and [*do* + *they* + V?] constructions:

(41) a. **Does** he need a tissue?
b. **Does** the kitty like milk?
c. **Does** the bears like milk?
d. **Do** the bear like**s** apple?
e. **Do they** like apples?
f. **Do they** need a tissue?

The same pattern was found in Round 5, when both *do* and *does* were sometimes used correctly in target contexts (42a-b), but *does* was used as the non-target form in questions with plural nominal subject:

(42) a. **Does** the bug want more milk?

b. **Do they** like apples?

c. **Does** the moon guy**s** want some more milk?

The child did not achieve full accuracy within the timeframe of the study, as *does* substitution errors were still found in the last round of data elicitation, and do was used correctly only with the subject they, indicating that there was no variability in the auxiliary and subject slots in the [*do*+*they*+V?] construction.

Data from SBST

Similarly to DNNC and TRRK, SBST's questions in Round 1 were marked with *is* in both singular and plural contexts (43a-b) or uninverted (43c):

(43) a. **Is** the kitty need a tissue?

b. **Is** the bears like milk?

c. The bears want some more juice?

In Round 2, the auxiliary form *do* was used as a question marker in virtually all, singular and plural, contexts:

(44) a. **Do** the kitty want more milk?

b. **Do** the kitty likes hamburgers?

c. **Do** they need a tissue?

d. **Do** the bears like milk?

SBST demonstrated a quick progress to productivity in DO-questions, as from Round 3 onwards, *do* and *does* were used correctly in all obligatory contexts, with a variety of pronominal and nominal subjects.

In sum, the children appeared to start off with the [*Is* + N + V?] construction with a lexically specific question marker *is.* Two children, proceeded to acquire a lexically specific construction [*Do*+*they*+V?] with a correct auxiliary

form, but with a fixed pronominal subject, indicating limited productivity. At the stage when do was used as the universal question marker with various subjects, [*Do* + N + V?] was used as a partially specified construction (with the fixed form *do* being the specified part). This construction showed a certain degree of abstractness because it was used with a variety of verbs and with a variety of subjects. However, it is clear that the children did not realize at this stage that the form of auxiliary DO should vary depending on the number of the noun that follows it. By Round 5, DNNC and SBST, but not TRRK, appeared to establish a productive [DO+N+V?] construction with variability in all slots.

8.3.8 Development of do-*support in declaratives with negation*

In this section, the progress in the use of auxiliary DO in constructions with negation will be discussed, using data from three children out of 9 (DVDC, DNNC, and TRRK). Again, we will track increasing schematicity in the development of the subject and the main verb slots in the earliest schemas and look for evidence for piecemeal acquisition, i.e. when different forms of the same auxiliary will come in as parts of separate constructions or at different times in the development.

Data from DVDC

First, consider the data from the Spanish-speaking child DVDC (age 6;03), who had had 8 months of exposure at the onset of the study. In Round 1, the majority of instances of *don't* occurred in the formula *I don't know*. All correct uses of *don't* occurred with the subject *I*. Whenever the subject was different, the child did not use *do*-support with negation:

(45) a. No **I don't** [/] # **we no** have.
b. And this one **no** eat.
c. That **no** go over there?

Since in DVDC's L1 Spanish negative forms of the verb were formed by adding negation marker *no* to the main verb, these errors could be the result of L1

transfer. Importantly, this pattern was never applied to first person contexts. The contracted form *don't* appeared to be memorized as a negation marker only with the pronoun *I*. In Rounds 2 and 3 (15–20 months of exposure), the form *don't* was used with subjects other than *I* (46a), and past tense marking on DO appeared, as the child started producing the contracted *didn't* (46b).

(46) a. But we **don't** get to play.
b. And then she **didn't** take Stitch.

From Round 3 on, *doesn't* was used correctly in third person contexts.

(47) And somebody my class **doesn't** know what zero plus zero means.

In Rounds 4 and 5 (26-32 MOE), *don't*, *doesn't* and *didn't* were used correctly with a variety of subjects:

(48) a. Except you **don't** have to web.
b. And I **didn't** know how to swim.
c. The first one &e &e it **doesn't** look neat.

It is possible that *didn't* was acquired earlier than *doesn't* because it has one form for all subject types. This order of acquisition also follows the frequency-based prediction, since *didn't* is more frequent in adult speech than *doesn't*.

Data from DNNC

The early stage of acquisition for DNNC, a Mandarin-speaking child (age 5;04, 9 MOE), was similar to the one of the Spanish-speaking child DVDC, as the negation form *don't* was used correctly in the first person contexts. In third person contexts, *don't* was applied as a negation marker instead of *doesn't*:

(49) a. Um &skr &skr uh I **don't** how to say! (omitted *know*)
b. Uh sometime I **don't** like food.
c. <He don't he don't> [/] he **don't** sleep here.
d. But [/] but he **don't** throw <in the in the> [/] in the # tree .

In Round 2 and Round 3 (14–20 months of exposure), past tense was correctly marked on *do*, but also sometimes doubly marked on *do* and the main verb:

(50) a. And [/] <and I> [/] and I **didn't** forgot.
b. I **didn't** know yet!

The contracted form *doesn't* was not produced until Round 4 (51a-b), and in Rounds 4 and 5 (28- 34 MOE) there were still occasional tense marking errors with the past tense form *didn't* (51c):

(51) a. And he **doesn't** know where I live.
b. And then Sandy **doesn't** know that he needs air.
c. I [/] I think <I don't> [//] I **didn't** really like um **saw** it.

Data from TRRK

A similar pattern of the development of negation was observed in TRRK, an Arabic-speaking child (age 4;02, 8 MOE). TRRK used *don't* correctly with the pronoun *I*, and misused *don't* as a negation marker in third person singular contexts. *Didn't* was also used correctly as a part tense marker early on, like in the other two children.

(52) a. Reem **don't** come play with me.
b. Yeah <and and> [/] and &do she **don't** beat me.
c. She [/] <she didn't> [/] she [/] she **didn't** go with me in my home.

Unlike the other children, TRRK substituted *don't* as a negation marker even in copula BE contexts:

(53) a. I **don't** a girl (= I am not a girl)
b. No I **don't** be doggy!

Again similarly to DNNC and DVDC, this child did not start using the contracted form *doesn't* until Round 3:

(54) a. But that camera **doesn't** hear me.
b. But she's water gun **doesn't** work.

The forms *doesn't* and *didn't* were used correctly from Round 3 onwards, with occasional tense marking mistakes, as in (55):

(55) Probably I **didn't saw** a clown.

In sum, the development of do-support in negation constructions appeared to follow similar stages in the three children, with *don't* acquired first as an unanalyzed negation marker. The form *didn't* was acquired at the same time or slightly later than *don't*, and inconsistency with past tense marking were found even at late stages of acquisition. Finally, *doesn't* was the last form to be acquired, following the frequency distribution pattern in adult speech given in Table 8.4 where *don't* was the most frequent form, followed by *didn't*, which was in turn followed by *doesn't*.

8.3.9 The development of constructions with embedded clauses

A very large proportion of the children's uses of *do* occurred in the formula *I don't know* used as an answer to questions and in the formula *Do you know (what)?* used to ask a question. These instances of *do* were excluded from the

analyses of accuracy. However, at later stages of acquisition, these formulas were used as parts of more complex schemas [*I don't know* + Clause] and [*Do you know* + Clause ?], and the development of such structures will be discussed in this section.

Diessel & Tomasello (2002) argued that children's earliest sentences with embedded clauses do not have a hierarchical structure, but rather consist of a linear combination of a formulaic main clause and a question simply attached to it. According to Diessel & Tomasello, embedded questions with uninverted word order are less frequent than main clause questions and therefore take longer to acquire. If children pay attention to frequent word combinations and also start off with the strategy of linear concatenation of a main clause and an inverted question, we expect them to overuse inverted word order in embedded questions. An investigation of the children's complex clauses revealed that this was indeed the case, as their early embedded questions were erroneously inverted:

(56) a. Um # I don't know who's that. (SBST, L1 Spanish, R1)
b. And do you know where is China? (DNNC, L1 Mandarin, R2)
c. I don't know how old is she. (DNNC, R2)
d. Mmm I don't know what's their name. (SBST, R2)
e. <I didn't> [/] <I didn't> [//] I don't know where's mine! (SMNS, L1 Spanish, R2)
f. I don't know what are those. No I just don't know <what> [/] what are <those> [/] those stuff. (JNNH, L1 Mandarin, R3)

Also confirming the prediction is the observation that the children seemed to invert the auxiliary erroneously when it was a part of a frequent combination, such as *where's*, *what's* or *who's*, as in the examples in (56)[12]. When the auxiliary

[12] It can be argued that the Spanish-speaking children used inverted word order in embedded clauses because it is grammatical in their L1. While L1 transfer might reinforce the trend in the Spanish group, it can not account for the same error in the Mandarin-speaking children (56b,c and f).

or the *wh*-words were less frequent, the auxiliary followed the subject rather than the *wh*-word, resulting in correct uninverted order:

(57) a. I'm [//] I don't know how many friends I will invite. (SBST, L1 Spanish, R2)
b. Do you know why we can't come in this door? (SBST, R3)
c. Do you know what school I'm going to? (DVDC, L1 Spanish, R3)
d. Uh I don't know what kind of meat that is. (DNNC, L1 Mandarin, R3)

Both correct and incorrect inversion patterns persisted in Rounds 3 and 4, demonstrating that children possibly were not distinguishing main and embedded clauses, but were rather guided by common word combinations.

(58) a. And I [/] I don't know what's that called. (MRSS, L1 Mandarin, R4)
b. Um I don't really know what it's called. (DNNC, L1 Mandarin, R4)

(59) a. I dunno what's a versity. (LLKC, L1 Arabic, R4)
b. But I don't know what are they in English. (LLKC, R4)
c. I dunno know what they are. (SMNS, L1 Spanish, R4)

Complex sentences such as (58a) above could be the result of two formulas being put together, 'I don't know' and 'what's that called?'. Similarly, sentence (59a) could be a result of concatenating two formulas, *I dunno* and *What's a N*?. However, correct uninverted clauses were also attested from Round 2 onwards:

(60) a. <I want> [//] I don't know what I am going to do today.
b. But other people don't understand what we're talking about.

The children appeared to converge on the target word order by Round 5.

8.3.10 Summary of the case studies

The development of constructions with BE. The trajectory of development of auxiliary BE constructions was similar in all the children. The earliest form to be acquired was the third person form *is*, whereas the plural form *are* appeared to come in as part of *they're + V-ing* construction, which constitutes evidence for piecemeal acquisition. The order of acquisition of BE forms appeared to follow the predictions based on frequency, since the third person form *is* was acquired first and remained the universal substitution form over many months of development. The form *are*, which is a lower-frequency form in adult speech, came in later and took longer than *is* to generalize over various subject types. Variability in the BE slot in the construction was observed from Round 3 on.

The question construction was acquired later than the declarative construction, and initially the children formed questions by changing the intonation of a declarative structure or using *is* as a universal question marker added to the beginning of a declarative sentence, indicating that the children's productivity with this structure emerged gradually as well. In other words, it was not the case that the abstract operation of inversion was acquired instantaneously. Productivity in the question construction, expressed in the variability of the BE and N slots in the [BE+N+V-*ing*] structure, was observed after Round 3.

The development of questions with DO. Several stages were observed in the development of *do*-support in questions. In Stage I, children did not use *do*-support and formed questions by changing the intonation of a declarative sentence. Alternatively, the form *is* was used as a default question marker at the beginning of the utterance, further supporting the importance of frequency in the choice of substitution forms. In Stage II, auxiliaries *do* and *does* appeared in questions, and *does* was supplied correctly with pronominal and nominal subjects, indicating some productivity of the question construction. However, *do* was often used as a substitution form in singular contexts, and *do* was used in a target-like

way only in the combination *Do they…?* The children possibly had not established the dependency between the number of subject and auxiliary form in the [*do/does* + N + V?] construction. Instead, their choice of auxiliary appeared to be guided by the familiarity of the form, since they chose the more frequent form *do* as the substitution form. Stage III corresponded to correct use of both *do* and *does* in most target contexts. However, at this stage *does* was sometimes used as the non-target form in questions with a plural nominal subject (e.g. *Does the bears* ...?), and this error was found in some children even in the last round of the study. Overall, these stages in development indicated that various forms of the same auxiliary were not acquired at the same time and that the productivity of constructions was emergent.

The development of negation with DO. The development of *do*-support in constructions with negation also appeared to follow similar stages in the children. The contracted form *don't* was acquired first as an unanalyzed negation marker. The form *didn't* was acquired at the same time or slightly later than *don't*, but errors with past tense marking were found even at late stages of acquisition. The errors involved past tense marked twice on the auxiliary DO and on the main verb (e.g. *I didn't saw*). On the one hand, it is an indicator of emergent productivity, as the children realized that the tense had to be marked in the structure. On the other hand, they didn't yet establish an abstract [N+DO+*not*+Vinfinitive] construction. The third person *doesn't* was the last form to be acquired, following the frequency distribution pattern in adult speech: *don't* was the most frequent form, followed by *didn't*, and *doesn't* was the least frequent form.

The development of constructions with embedded clauses. Confirming the prediction of a frequency-based approach, the study found that the children erroneously inverted the auxiliary in embedded clauses when it was a part of a frequent contractible combination, such as *where's*, *what's* or *who's* (e.g. *I don't know what's that called*). Both correct and incorrect patterns persisted until Round 4, demonstrating that children still did not distinguish main and embedded

clauses, but were rather guided by common word combinations. It was concluded that complex sentences with embedded clauses were the result of two unproductive constructions put together, e.g. *I don't know* and *What's that called?*, or the result of concatenating an unproductive construction with a partially variable construction, e.g. *I dunno* and *What's a N?*. However, correct uninverted clauses were also attested from Round 2 onwards, alternating with clauses with incorrect word order. The children appeared to converge on the target word order by Round 5.

8.4 Summary

The aim of the research reported in this chapter was to determine the pattern of auxiliary acquisition by analyzing longitudinal data collected from 9 children over a two-year period in their course of development. More specifically, the research was aimed at establishing whether children show differences in the accuracy with different forms of BE and DO in declaratives and questions, whether the accuracy reflects the frequency distribution of auxiliaries in adult speech, and whether the development of auxiliary constructions is piecemeal.

(1) Does the order in which children acquire specific auxiliary forms reflect their relative frequencies in the language children hear? (a) High-frequency forms were expected to have higher rates of provision; (b) High-frequency auxiliary forms were expected to be substituted for lower frequency forms. The data supported these predictions. Children were more accurate with the high-frequency auxiliary *is*, used it as a substitution form for the plural form *are*, and as a substitution for *do* and *does* in yes/no questions. Among the DO forms, *do* was initially used as the substitution form for *does* and *did*, as expected, but only at the early stages of acquisition. In elicitation tasks at later stages of acquisition, children started to overuse the form *does* in plural contexts, perhaps signaling emerging productivity, as the children acquired a variety of forms and began to use them creatively. The auxiliary DO demonstrated very different acquisition trajectories in negative structures and yes/no questions, indicating that there was no generalization across *do*-support in constructions with negation and

in questions, at least at the early stages of acquisition. This could partly be due to input frequency as well, since DO with negation is a more frequent structure than DO in yes-no questions (Table 8.4).

The issue regarding the process of emerging productivity with auxiliary verbs was addressed in two research questions:

(2) What are the effects of the subject type? Children's proportional use of different auxiliary forms was predicted to be influenced by an interaction between type and token frequency: (a) High rates of auxiliary provision were expected in combinations with high-frequency pronouns in the form of pronoun+auxiliary chunks; (b) Auxiliary suppliance for NP subject forms was expected to be lower than for fixed pronoun+auxiliary combinations. These predictions were supported, as the children were more accurate with auxiliary BE in pronoun subject contexts than in NP subject contexts.

(3) What is the process of developing productivity in auxiliary constructions? (a) Increasing schematicity could be seen in the development of the subject slot in the earliest schemas. Most schemas appeared initially with a fixed subject (e.g. *Do they* + V? and *They are +Ving* were the only contexts in which *do* and *are* were used correctly). Over time, subject variation appeared in the initially item-specific schemas. Also with respect to prediction (a), some support was found in the children's use of the contracted negative form *don't*. Some children indeed appeared to use it exclusively as a negation marker in first person contexts *I don't* + V. (b) Different forms of the same auxiliary came in at different stages of development. Children used *don't* with a variety of subjects, applying it as a universal negation marker (e.g. *I don't*, *he don't*, N + *don't*), with forms *didn't* and *doesn't* appearing months later than *don't*. Thus, it took them some time to establish generalizations over these forms of the auxiliary DO. The development of DO forms in questions and constructions with negation was different, which was reflected in different accuracy and substitution patterns in these two constructions. Finally, emergent productivity was observed in the development of complex constructions consisting of [*I don't know* + Clause] and

[*Do you know* + Clause] parts in combination with an embedded question, which initially showed signs of linear construction-building.

In sum, this chapter investigated children's acquisition of auxiliary verbs BE and DO in declaratives and questions. Forms produced spontaneously were compared with structures produced in an elicitation task, and the children's accuracy and error patterns were investigated over two years of development. The data suggest that different forms of auxiliaries were acquired separately and followed different paths of development. The children demonstrated some generalization over various BE forms, resulting in high accuracy scores in different person and number contexts in both spontaneous and elicited speech. Forms of auxiliary DO in questions and declaratives showed a lack of generalization across various kinds of *do*-support. The use of auxiliaries was influenced by input frequencies, as more frequent forms were supplied more accurately and were used as substitution forms for less frequent forms. Also, the children's suppliance of auxiliaries was influenced by not just the frequency of the individual auxiliary forms, but by frequencies of subject+auxiliary combinations. The children were more accurate supplying auxiliaries with pronominal subjects than with nominal subjects. Taken together, these findings indicate that the L2 children's acquisition of auxiliary constructions was piecemeal, as predicted by the usage-based model.

CHAPTER 9. CONCLUSION

9.1 Introduction

The analyses reported in this thesis were aimed at exploring several theoretical issues in child L2 acquisition that are either controversial or have not received attention in previous work. These issues included L1 transfer, the role of input frequency and the process of the development of productive structures and abstract knowledge in L2. These issues were considered in the light of two major theoretical approaches to language acquisition, namely the constructivist approach (in particular, the usage-based model), and the generative approach (in particular, the Full Transfer/Full Access model and the feature assembly model).

The data that were used to assess these models came from two studies of children learning English as a L2, a longitudinal and a cross-sectional study. The children came from a variety of L1 backgrounds, namely Arabic, Cantonese and Mandarin Chinese, Hindi, Urdu or Punjabi, and Spanish. Both spontaneous and elicited speech was used, and the linguistic phenomena chosen for the analysis were the article system and the auxiliary system of English. This chapter summarizes the findings of the studies with respect to each issue in question and discusses the implications for the theoretical models.

9.2 L1 transfer in articles and auxiliaries

Two versions of a UG-based approach to L2 acquisition were tested, the Full Transfer/Full Access (FT/FA) model and the feature assembly model. Predictions of the FT/FA model regarding the transfer of abstract functional structure of noun phrases were confirmed in both studies. Recall that both studies found that the L1 typology could indeed facilitate the acquisition of the article system, but only for the general awareness of the obligatoriness of articles. The particular *type* of the article system in the background of L2 learners did not seem to have an effect. Particular differences between L1s such as Spanish and Arabic did not seem to matter for the development of the article system in child L2 learners of English. In terms of the feature-assembly approach, the children did

not transfer particular feature-morpheme mappings from L1, but rather transferred a general knowledge of an abstract functional layer in the NP structure, which is more compatible with the FT/FA model.

In the study of L1 transfer effects in the acquisition of auxiliary verbs in Chapter 6, similar conclusions were made. With respect to accuracy with auxiliary BE, the only significant L1 effect apparent in the data was the difference in accurate suppliance of BE in declarative sentences. The significant effect on accuracy was due to the low suppliance of the auxiliary in the Chinese L1 group, while the other three groups were equally producing correct declaratives with BE. There were no significant differences in the accuracy among the Arabic, Spanish, and Hindi L1 groups. These findings support the predictions of the FT/FA model, indicating that the presence of the abstract categories of Tense and Agreement facilitated acquisition, rather than more specific similarities in the grouping of features into lexical items predicted by the feature assembly account. Both FT/FA and feature assembly models predicted that all children would have difficulty with DO-questions because it is an idiosyncratic structure of English, resulting in lower accuracy with DO than with BE. This prediction was borne out in the data, as the statistical tests revealed no significant differences among mean accuracy rates with DO.

Putting together the results of the studies discussed in Chapters 5 and 6, what is transferred from L1 in L2 acquisition of the article and auxiliary systems? The findings in both studies provided evidence for the knowledge of abstract categories in the nominal and verbal domain in child L2 learners. Two versions of a UG-based transfer model were tested, the Full Transfer/Full Access and the feature assembly. Predictions of the Full Transfer model regarding the transfer of abstract functional structure were confirmed both in the two studies of article acquisition and in the study of auxiliary acquisition. An interesting generalization emerging from the findings of all the studies taken together is that the L1 typology can indeed facilitate the acquisition of the functional structure of noun phrases and verb phrases, but only extends as far as the general awareness of the obligatoriness of the functional morpheme. The particular *type* of the article

system or auxiliary system in the background of L2 learners did not have an effect, as these performed similarly regardless of whether the feature-lexical item mappings were parallel in their L1 and L2 English.

Another common finding with regard to L1 transfer in the nominal and verbal domain in child L2 English was the limited nature of L1 effects, as group differences were observed only in the first 1.5 years of acquisition. Compared with previous work on adult L2 acquisition the role of L1 was less pronounced in the children's developmental patterns and rates of acquisition, compared with adult L2. For instance, previous studies reported very few article errors in L2 English of adult learners whose L1 article systems were similar to English, such as French (Sarko, 2008), Greek (Hawkins et al., 2006), or Spanish (Ionin et al., 2008; Snape et al., 2006). A study of adult speakers of Arabic (Sarko, 2008) found more specific L1-influenced errors in their article choice, as they omitted indefinite articles but not definite articles. As for the rate of acquisition, children demonstrated significant improvement even at the early stage and converged on the target system faster than adults studied in previous work on L2 acquisition of articles (Chapter 5, Section 5.3.4). The short-lived nature of L1 transfer in child L2 has already been mentioned in previous studies. For instance, Selinker & Lakshmanan (1992) suggested that children may be more successful than adults in their reanalysis on the target language input, thus overriding L1 transfer faster. Meisel (2008) explained the difference between adult and child L2 learners by suggesting that whereas adult L2 learners have a complete and stabilized L1 system at the onset, L1 knowledge in child L2 learners is still incomplete and probably unstable.

In addition, in the nominal and verbal domains there were developmental trends that might be frequency- rather than L1-based. These trends were explored further in light of the usage-based framework in Chapters 7 and 8, and the findings of these studies and the implications for the interaction of abstract and lexically-specific knowledge in L2 children will be discussed in the next section.

9.3. The role of frequency

An important finding in the study of the children's acquisition of articles in Chapter 5 was that the definite article was 'easier' to learn and that it was erroneously used as the substitution form in indefinite contexts. Differential difficulty by article type does not follow directly from the FT/FA predictions. The feature-assembly approach can be applied to account for the difficulty that learners in both studies had with the mastery of the indefinite article, regardless of their L1 background. In a discussion of article acquisition in an adult L2 learner, Lardiere (2004: 335) suggested that differences in accuracy in indefinite and definite contexts could be due to the fact that "definite articles in English need not take number and the count/mass distinction into account, which makes them less featurally complex than indefinites in at least one respect". A similar conclusion can be drawn from Hawkins et al.'s (2006: 20) discussion of featural contexts of insertion for articles: *a* is inserted in [D, -definite, +singular] contexts, *the* is inserted in [D, +definite] contexts. Thus, in order to use *a* appropriately, learners have to identify the feature [singular] as relevant for the insertion of this article.

However, several findings point to the conclusion that the featural complexity approach is not sufficient to explain differential accuracy. First, most errors of *the* substitution occurred when the target indefinite noun was the subject of the sentence, while existential constructions and object positions were the contexts where the indefinite article was often used by the children correctly, suggesting the influence of the syntactic context on the children's choice of articles. This does not necessarily imply, however, that they used *a* without understanding its semantics, because *a* was sometimes correctly used in subject positions from Round 1 on, and at the same time *a* was not overused in definite contexts. But the fact that the children were more accurate supplying the indefinite articles in some constructions than in others is a strong piece of evidence for piecemeal acquisition of the structure of the noun phrase and it cannot be accommodated by the generative approaches.

It is possible to account for differences in difficulty without appealing to features available from UG. From a usage-based point of view, reasons for

accuracy and overuse pattern can be determined by the distribution of articles in the input. First of all, *the* is more frequent in speech than *a*. The fact that the definite article is more frequent than the indefinite article could be a factor contributing to the definite article becoming the 'default' form and being acquired first. A closer look at the distribution of articles in the input revealed that *the* is more frequent with subject nouns, whereas *a* is more frequent in object contexts, following a verb or a preposition.

Frequency effects were found in the verbal domain as well. High-frequency forms of auxiliary verbs had higher rates of provision and high-frequency forms were also substituted for lower frequency forms, as predicted by the usage-based model. In the case of BE, children were more accurate with the high-frequency auxiliary *is*, used it as a substitution form for the plural form *are*, and as a substitution for *do* and *does* in yes/no questions. As for DO forms, the higher-frequency *do* was initially used as the substitution form for *does* and *did* at the early stages of acquisition.

A potential argument against frequency would be that regardless of differences in frequency of *the* and *a* or *is* and *are*, all these words are extremely frequent in English, and yet L2 learners demonstrate protracted phases of omission or lack of in consistency in choosing the right forms. In other words, while saying that the overall frequency of *a* is smaller than that of *the* we need to remember that *a* is still one of the most frequent words in the language and thus the comparison of the overall frequencies of articles per se may not be very informative. This is precisely why the finer-grained analysis in Chapter 7 was necessary, comparing of article use in various lexical positions and with various lexical items. It was necessary to look at contexts with lower absolute frequencies to establish that (1) the overall dominance of *the* over *a* is present at the level of lexically specific constructions as well, and (2) that at this level it is possible to find the weakening of the dominance (with *a* being more frequent than *the*).

Furthermore, frequency is not the only factor that plays a role in acquisition sequences in the usage-based approach. Previous studies also named factors such as acoustic salience of morphemes and complexity of semantic

function-form mapping in constructions that morphemes are part of (Behrens, 2009; Ellis & Larsen-Freeman, 2006, Lieven & Tomasello, 2008). For instance, articles are functional words that are difficult to perceive on the basis of purely acoustic evidence. However, in comparison with the indefinite article *a*, the definite article *the* can be more acoustically salient, which can be an additional explanation for why learners use *the* as the substitution form. Semantics can also be a factor because the semantic function that articles are mapped to can only be established in context with their collocates. Thus, input frequency is by no means the only factor that influences language development in the usage-based view. Factors such as frequency, saliency, and semantic and structural complexity of constructions together can account for patterns in L2 acquisition without reference to abstract notions such as the functional category Determiner or abstract features such as [±definite]. The weakness of the usage-based analysis that we should be aware of is that it makes no specific predictions for the additional factors (i.e. semantics, acoustics, etc.) and only refers to them post-hoc when the findings cannot be fully accounted for by input frequency. This weakness has been mentioned in usage-based papers reporting findings that are not predicted by input frequency (Ambridge & Rowland, 2009; Theakston & Rowland, 2009a,b), as well as in generativist papers (e.g. Anderssen & Westergaard, 2010). It should also be pointed out, however, that these other factors, which are not considered in the present study, are theory-neutral and can be cited by both generativists and emergentists. In the present study, the emergentist approach is favoured on the basis of such distinguishing factors as frequency effects and piecemeal acquisition, to be discussed in the following section.

9.4 Piecemeal acquisition of constructions with articles and auxiliaries

High rates of auxiliary provision were found in combinations with high-frequency pronouns in the form of pronoun+auxiliary chunks. Auxiliary suppliance for NP subject forms was lower than for fixed pronoun+auxiliary combinations. The fact that the children were more accurate with auxiliary BE in pronoun subject contexts than in NP subject contexts indicates that inflected

forms of auxiliaries come in as parts of lexically specific constructions, such as *you're V-ing* or *it's V-ing*. Individual analysis of children's development revealed that months passed between the times the children started using *is* and *are* in elicitation tasks, and the first uses of *are* were limited to *they're* combination. There was a delay between the appearance of first instances of the use of certain forms in pronoun-auxiliary combinations and their use in constructions with a variable subject such as *NP's V-ing*. These findings indicate that the children were not operating with abstract categories such as Tense, Agreement, or Aspect, and did not have the knowledge of the abstract clause structure in general. Rather, they appeared to proceed in building the constructions in an item-by item fashion, gradually increasing their productivity.

The acquisition of DO also provided evidence for piecemeal acquisition of constructions. The contracted form *don't* was acquired first as an unanalyzed negation marker. The form *didn't* was acquired at the same time or slightly later than *don't*, but errors with past tense marking and number agreement were found even at late stages of acquisition (at the end of the two-year longitudinal study). These patterns cannot be accounted for in terms of categories like Tense and Agreement, since different auxiliaries, and even different forms of the same auxiliaries, demonstrated different routes of acquisition. These findings are similar to the findings in the nominal domain, where articles *a* and *the* appeared to be acquired initially as parts of different constructions, with accurate uses of *a* being initially tied to lexically specific combinations such as *there was a NP*.

Finally, limited productivity was found not only at the word and phrase level, but also at the clause level. The study of the development of complex clauses (Section 8.3.9) revealed that sentences with embedded clauses do not have a hierarchical structure, but rather consist of a linear combination of a formulaic main clause and a question attached to it, resulting in non-target inverted word order in the embedded clause (e.g. *I dunno what's what*). Again, children seem to anchor their early constructions with a lexically specified combination such as *I dunno* and gradually increase the variability of the elements of the combination, as well combine it with other constructions to make a more complex construction.

Taken together, the patterns of acquisition in verb phrases, noun phrases, and clause structure demonstrated that the productivity of constructions that involved articles and auxiliaries was emergent and that the process of acquisition was piecemeal.

9.5 The usage-based model vs. the generative model

The usage-based model successfully predicted the patterns in child L2 acquisition reported in Chapters 7 and 8. While the generative model can accommodate some input frequency effects, the usage-based approach is better at explaining differences in accuracy with different inflected forms of auxiliaries, differences in accuracy with definite and indefinite articles, and protracted stages of low productivity. The usage-based model does not make a distinction between abstract core syntax and other levels of grammar. Rather, all linguistic phenomena are considered to be form-function correspondences. This flexibility in what is considered word-level and what is considered category-level phenomena is useful in explaining why some forms of the same general category take longer to acquire and why some constructions are more difficult than others.

It has been acknowledged that the usage-based approach is somewhat vague with respect to how frequency interacts with other usage factors in acquisition (see Section 9.3). It is also necessary to point out another limitation of the usage-based analyses given in the literature and in the present study. The analyses in the study were based on production data and did not include tests of the children's comprehension of grammar. The generative model could be used to compare these two types of data since this approach states a clear distinction between the learners' competence and performance and expects that performance can 'lag behind' competence in children. The usage-based model would not be able to formulate predictions for the comprehension/production comparison simply because it is not clear how this approach would distinguish the two notions.

However, in the production data several types of evidence were found that indicate that the usage-based approach is better placed to explain the trends that

emerged in the acquisition of articles and auxiliaries and various constructions that they are part of. First, patterns of piecemeal learning (e.g. the gradual growth of the inventory of inflected forms of BE) challenged the claim that children's early performance is adult-like in the sense that it relies on the pre-given knowledge of abstract categories (e.g. Tense and Agreement). This type of evidence is taken to prove that children's early categories are not abstract and category-general, but rather lexically specific.

Second, comparisons of distributions of auxiliaries and articles in adults' speech and children's speech demonstrated that children's early usage is influenced by input frequency, i.e. it can be related to the patterns that children often hear in adult speech. It is necessary to go back to the discussion of the role of frequency in generative and usage-based accounts initiated in Chapter 2 (Sections 2.2-2.3). Lexical input frequency does have a place to influence order of acquisition in the generative view as well because regardless of the theoretical approach, language acquisition is the result of exposure to input. Even in the generative view, acquisition does involve learning language-specific inflected forms. However, external conditions such as frequencies of various forms are considered necessary only to activating the knowledge of abstract categories that are *already instantiated* in the given language. Thus, the generativist view can accommodate the fact that the more frequent inflected form is acquired first (e.g. *is* vs. *are*). However, once the forms are used by the child, the generative account would not predict context-specific use (e.g. using *are* accurately with the subject *they* but not with plural NPs). Thus, the generative view does consider input frequency as a factor, but it can accommodate only some of the observed frequency effects.

Finally, analyses of the productivity of complex sentences involving embedded clauses (Section 8.3.9) demonstrated that the children's learning was piecemeal at the sentence level as well and was better explained by linear concatenation strategies with gradual generalization over variable slots in constructions than by operations involving abstract categories.

9.6 The usage-based model and L1 transfer

We need to return to the issue of L1 transfer in UG-based and usage-based models of L2 acquisition, as it was mentioned earlier in the discussion of theoretical approaches in Chapter 2 (Section 2.3.4) that the transfer of abstract functional structure has not yet been formulated from a usage-based perspective. In the case of auxiliary BE for instance, the facilitation effect that was found in Arabic, Hindi and Spanish learners (Ch. 6, Section 6.3) appeared to be based on a very general structural property of typologically different L1s, which would be hard to accommodate under the hypothesis that L2 learners follow a developmental path in acquisition solely based on input. L1 transfer effects in child L2 learners provide evidence for the ability of these children to generalize across constructions and morphemes (auxiliaries and articles) and use this knowledge to give them a head start in L2.

How can these findings be explained in terms of the usage-based model? Given that transfer from one language to another can be considered a type of overgeneralization, and that overgeneralizations in acquisition occur when children discover commonalities (structural and/or semantic) across several elements of language, we can conclude that children had the knowledge of some abstract constructions already at the earliest stage of L2 acquisition. There have to be patterns that are so similar in the two languages that the child cannot help comparing them and drawing them together. However, the patterns in question have to be abstract enough that they do not rely on lexically specific information, and thus are in principle transferable across two distinct sets of lexical items (e.g. Hindi, unlike English, has two distinct auxiliaries marking tense and aspect, and the order of elements in the verb phrase in the two languages is completely different – see Chapter 3, Section 3.5). This can explain why we do not find transfer of lexically specific patterns, but rather transfer of broad abstract patterns. Then how can we formalize the transfer of language structures from L1 to L2 in terms of the usage-based model? In order to do this, it is necessary to go back to the definition of a construction and the schematic representation of the syntax-

semantic mapping suggested by Goldberg (2006: 20), using the example of the ditransitive construction.

(1) Semantics: Intend-CAUSE-RECEIVE (agent – recipient – theme)

Syntax: verb Subj Obj1 Obj2

On the semantic level, the meaning of each of the element is described using a label for convenience. Importantly, this label is not a universal semantic role but simply a word generalizing typical meaning of this slot in the construction. Importantly, the surface form of a construction need not specify a particular word order, which was the case for constructions transferred from children's L1s.

First consider the case of the NP construction. In order to analyze the syntactic level of the construction, we can apply Goldberg's approach and represent the NP construction transferred from L1s to L2 English as follows:

(2) NP construction: {(Article) (Adjective) Noun}[13]

Speakers of Arabic and Spanish (and possibly Romainan, since there was one Romanian-speaking child in the [+article] group) transferred the NP construction from their L1 with the article slot present. This slot was more abstract than an array of specific morphemes from L1 because, for example, the Spanish-speaking children did not produce phrases like *el elephant* or *el giraffe* with article forms directly transferred from Spanish. However, this slot was more specific than say a general category of Determiner. It is necessary to use the term *article* for this slot in the NP construction because the presence of determiners such as demonstrative pronouns or classifiers in L1s did not help Hindi- or Chinese-speaking children with learning a new type of determiner. The children who transferred the NP construction given in (1) from their L1s demonstrated awareness of the fact that the slot before the noun in the NP must be occupied by a functional element and

[13] Curly braces indicate that the word order is not fixed. Parentheses indicate optional elements

that this functional element contains some information about the newness of the entity represented by the noun. This knowledge was demonstrated when the children made substitution errors in indefinite contexts but not definite contexts, i.e. their use of articles was not simply random.

As for the order of elements in the NP construction transferred from L1s, the children did not transfer it with a fixed word order. For instance, in Spanish and Hindi adjectives follow nouns, but speakers of these languages did not make word order mistakes in English NPs. The construction that they applied to English did not have a fixed word order, otherwise they would have produced English nouns followed by adjectives rather than preceded by them. Even the order of the article and the noun was not fixed because in Romanian, the definite article follows the noun, but the Romanian-speaking child in the [+ article] group did not make errors with the order of *the* and noun in English. Thus, in the children's knowledge of NP structure, there was abstraction at this level as well.

Next we turn to the constructions involving auxiliary BE. L1 structure that facilitated the acquisition of this construction in L2 English of Arabic, Spanish, and Hindi children can be formulated as the following construction:

(2) Progressive construction: {Subj AuxV lexical V_{prog}}

First of all, just like in the case of the NP construction, the children appeared to be aware of the semantics of this construction, since it was used appropriately to describe continuous action. With respect to the syntactic form, just like in the case of the NP construction, some abstractness and some specificity was demonstrated in the elements of construction. Just like in the case of the NP construction, the children clearly did not transfer specific lexical items from their L1 as part of the construction (e.g. the auxiliary *está* 'is' or the verb ending *–ndo* '-ing' in Spanish). Furthermore, the construction that the children transferred from their L1s did not have word order specified. This conclusion is based on the observation that Hindi-speaking children did not make word order errors despite the fact that in Hindi auxiliary verbs always occupy sentence-final position. In the

Arabic progressive construction, the auxiliary occupies sentence-initial position, followed by the subject, and it is used only in the past tense. The fact that these children did not transfer these specific properties of the construction indicates that their progressive construction was abstract enough to not specify the order of the three slots. At the same time they demonstrated awareness of the fact that the construction requires a progressive form of the lexical verb rather than, say, a bare infinitive, because the children started using the *–ing* form of the verb correctly from the onset.

Recall that all children had difficulty acquiring the inverted word order in questions with BE. It seems to be the case that what Arabic-, Spanish-, and Hindi-speaking children transferred from their L1 was simply the declarative structure in (3). The BE question construction with inverted word order given in (4) could not be transferred from L1 and thus was challenging for all L1 groups.

(3) {Subj AuxV lexical V_{prog}} + question intonation

(4) AuxV Subj lexical V_{prog}

An important observation is that the children for whom the English word order in all types of sentences was new (i.e. Arabic and Hindi) and the children for whom auxiliary verbs were new altogether (i.e. Mandarin and Cantonese) all had more difficulty with question word order than with declarative word order (Figures 6.1 and 6.2 in Chapter 6). Partly this difference can be due to the fact that English does have the option of asking a question simply by changing the intonation, and partly it can be due to frequency, since declaratives are more frequent than inverted questions. In other words, when the children's L1 could not facilitate the acquisition of the structure in any way, there were other factors specific to English that influenced the acquisition order. It could also be the case that the children made more L1 errors in the structures that are less frequent in L2. For instance, if *He's*+Ving is more frequent in English than *NP're*+Ving, the children could have the former structure stored as a partially specified construction and use an L1-based structure (e.g. NP with a bare main verb) instead of the latter structure.

Thus, there is an interplay of L1 transfer and frequency factors in child L2 acquisition.

Having transferred abstract constructions such as (2) and (3) from the L1, the children were initially conservative in their acquisition of the specific lexical items that go in the slots of these constructions. In other words, even though the children demonstrated a substantial degree of abstractness in their syntactic structures given in (2) and (3), the process of filling the slots in these structures was the process of piecemeal learning. The children added only slowly to the forms that each auxiliary verb can take or to the subject forms that can co-occur with auxiliaries rather than plugging all similar items in the slots right away. Importantly, this item-by-item acquisition process was the same in all children regardless of their L1 background. This process was not influenced by the fact that some children initially got a head start, being able to transfer the English-like construction form for noun phrases and verb phrases from their L1.

In sum, the usage-based model still can account for the process of acquisition in L2 children, but with modifications concerning the availability of abstract constructions already at the onset of acquisition. L1 acquisition is seen as a gradual process of generalization from memorized chunks to partially lexically specific constructions to abstraction across constructions and emergence of higher order relations among groups of abstract constructions. However, recent studies of L1 acquisition have already started to modify this view of the acquisition process (Abbot-Smith & Behrens, 2006; Lieven & Tomasello, 2008: 182-183; Siebenborn, 2010). They point out that the process of generalization is not necessarily gradual since some constructions are clearly acquired earlier than others and explore how previously learned structures can facilitate or inhibit acquisition of a particular target construction – within one language. This is even more obvious in L2 acquisition, where the processes of item-by-item learning from input and generalizations across constructions appear to be intertwined already at the early stage, possibly due to the fact that L2 learners are older and thus have bigger working memory and better abilities to create generalizations, discern analogies and combine structures.

BIBLIOGRAPHY

Abbot-Smith, K. & Behrens, H. (2006). How known constructions influence the acquisition of other constructions: The German passive and future constructions. *Cognitive Science, 30*, 995–1026.

Adger, D. (2003). *Core syntax: A minimalist approach*. Oxford University Press.

Ambridge, B. & Rowland, C. (2009). Predicting children's errors with negative questions: Testing a schema-combination account. *Cognitive Linguistics, 20*, 225–266.

Ambridge, B., Rowland, C., Theakston, C. & Tomasello, M. (2006). Comparing different accounts of inversion errors in children's non-subject wh-questions: 'What experimental data can tell us?'. *Journal of Child Language*, *33*, 519–557.

Anderssen, M. & Westergaard, M. (2010). On the impact of frequency in the acquisition of variable word order. In K. Franich, K. M. Iserman, & L. L. Keil (Eds.), *Proceedings of the 34th annual Boston University Conference on Language Development, Vol. 1* (pp. 1-10). Somerville, MA: Cascadilla Press.

Anderssen, M. & Westergaard, M. (in press). Frequency and economy in the acquisition of variable word order. *Lingua, Special Issue*.

Bateson, M. C. (1967). *Arabic language handbook*. Washington, DC: Center for Applied Linguistics.

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47,* 383–411.

Bhatia, T. K. (1993). *Punjabi: A cognitive-descriptive grammar*. London and New York: Routledge.

Bickerton, D. (1981). *Roots of language*. Ann Arbor: Karoma Press.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Butt, J. & C. Benjamin (1994). *A new reference grammar of modern Spanish.* London: Arnold.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10,* 425–455.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University Press.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson and N. Ellis (Eds). *Handbook of cognitive linguistics and second language acquisition* (p. 216–236). New York: Routledge.

Bybee, J. & Hopper, P. (Eds.), (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Bybee, J. & Scheibman, J. (1999). The effect of usage on degree of constituency: The reduction of *don't* in American English. *Linguistics, 37*, 575–596.

Cameron-Faulkner, T., Lieven, E. & Tomasello, M. (2003). A construction based analysis of child-directed speech. *Cognitive Science, 27*, 843–873.

Chen, P. (2004). Identifiability and definiteness in Chinese. *Linguistics, 42*, 1129–1184.

Cheng, L. L.-S. & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry, 30,* 509–542.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chondrogianni, V. (2008). Comparing child and adult L2 acquisition of the Greek DP: Effects of age and construction. In B. Haznedar & E. Gavruseva (Eds.), *Current trends in child second language acquisition* (pp. 97–144). Amsterdam: John Benjamins.

Clahsen, H. (Ed.), (1996). *Generative perspectives on language acquisition: Empirical findings, theoretical considerations, and cross-linguistic comparisons*. Amsterdam: Benjamins.

Croft, W. (2001). *Radical construction grammar*. Oxford: Oxford University Press.

Dabrowska, E. (2000). From formula to schema: The acquisition of English questions. *Cognitive Linguistics, 11,* 83–102.

Dabrowska, E. & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics, 16,* 437–474.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present*. Available online at http://www.americancorpus.org.

De Cat, C. (in press). Information tracking and encoding in early L1: linguistic competence vs. cognitive limitations. *Journal of Child Language*.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22,* 499–533.

Diessel, H. & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A usage-based approach to the development of grammatical constructions. *Cognitive Linguistics, 12,* 97–141.

Dulay, H. & Burt, M. (1973). Should we teach children syntax? *Language Learning, 24,* 245–258.

Dulay, H. & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning, 24,* 37–53.

Ellis, N. (2002). Frequency effects in second language processing. *Studies in Second Language Acquisition, 24*, 143–188.

Ellis, N. (2006). Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27, 1–31.

Ellis, N. & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics. Introduction to the Special Issue. *Applied Linguistics, 27*, 558–589.

Gass, S. & Mackey, A. (2002). Frequency effects and second language acquisition: A complex picture? *Studies in Second Language Acquisition, 24*, 249–260.

Gathercole, V. C. Mueller (2007). Miami and North Wales, so far and yet so near: A constructivist account of morphosyntactic development in bilingual children. *The International Journal of Bilingual Education and Bilingualism, 10,* 224–247.

Gathercole, V. C. Mueller, Sebastián, E., & Soto, P. (1999). The early acquisition of Spanish verbal morphology: Across-the-board or piecemeal knowledge? *International Journal of Bilingualism, 3,* 133–182.

Gavruseva, E. (2008). The development of copula and auxiliary *be* and overgeneration of *be* in child L2 English. In B. Haznedar & E. Gavruseva (Eds.), *Current trends in child second language acquisition: A generative perspective* (pp. 145–177). Amsterdam: John Benjamins.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Grondin, N. & White, L. (1996). Functional categories in child L2 acquisition of French. *Language Acquisition, 5*, 1–34.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24,* 287–298.

Hatch, E. (1972). Some studies in language learning. *UCLA Workpapers in Teaching English as a Second Language, 6,* 29–36.

Hawkins, R. (2001). *Second language syntax*. Oxford: Blackwell.

Hawkins, R., Al-Eid, S., Almahboob, I., Athanasopoulos, P., Chaengchenkit, R., Hu, J., et al., (2006). Accounting for English article interpretation by L2 speakers. In S. Foster-Cohen, M. Medved Krajnovic & J. Mihaljević Djigunović (Eds.), *EUROSLA Yearbook: Vol. 6* (pp. 7–25). Amsterdam: John Benjamins.

Hawkins, R. & Chan, Y.-H. C. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research, 13,* 187–226.

Hawkins, R. & Hattori, H. (2006). Interpretation of English multiple wh-question by Japanese speakers: A missing uninterpretable feature account. *Second Language Research, 22*, 269–301.

Haznedar, B. (1997). L2 acquisition by a Turkish-speaking child: Evidence for L1 influence. In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of*

the 21st Annual Boston University Conference on Language Development (pp. 245–256). Sommerville, MA: Cascadilla Press.

Haznedar, B. (2001). The acquisition of the IP system in child L2 English. *Studies in Second Language Acquisition, 23,* 1–39.

Haznedar, B. (2007). The acquisition of tense–aspect in child second language English. *Second Language Research, 23,* 383–417.

Haznedar, B. & Gavruseva, E. (Eds.), (2008). *Current trends in child second language acquisition: A generative perspective*. Amsterdam: John Benjamins.

Haznedar, B. & Schwartz, B. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 257–268). Sommerville, MA: Cascadilla Press.

Hegarty, M. (2005). *A feature-based syntax of functional categories: The structure, acquisition, and specific impairment of functional systems*. The Hague: Mouton de Gruyter.

Herschensohn, J. (2000). *The second time around: Minimalism and L2 acquisition*. Amsterdam: John Benjamins.

Hu, J., Pan, H. & Xu, L. (2001). Is there a finite vs. nonfinite distinction in Chinese? *Linguistics, 39,* 1117–1148.

Huebner, T. (1985). System and variability in interlanguage syntax. *Language Learning, 35,* 141–163.

Ionin, T. (2003). *Article semantics in second language acquisition*. Doctoral dissertation, MIT. MIT Working Papers in Linguistics.

Ionin, T. (2008). Progressive aspect in child L2-English. In B. Haznedar & E. Gavruseva (Eds.), *Current trends in child second language acquisition: A generative perspective* (pp. 17–55) Amsterdam: John Benjamins.

Ionin, T. & Wexler, K. (2002). Why is 'is' easier than '–s'?: Acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research, 18,* 95–136.

Ionin, T. & Wexler, K. (2003). The certain uses of *the* in L2-English. In J. M. Liceras et al. (Eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference*, (p. 150 – 160). Cascadilla Press.

Ionin, T., Ko, H. & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition, 12,* 3–69.

Ionin, T., Zubizarreta, M. L. & Maldonado, S. B. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua, 118*, 554–576.

Ionin, T., Zubizarreta, M. L. & Philippov, V. (2009). Acquisition of article semantics by child and adult L2-English learners. *Bilingualism: Language and Cognition, 12,* 337–361.

Jia, G. & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research, 50,* 1280–1299.

Johnson, J. & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational age on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Kachru, Y. (2006). *Hindi*. Amsterdam: John Benjamins.

Klima, E. & Bellugi, U. (1966). Syntactic Regularities in the Speech of Children. In J. Lyons & R. Wales (Eds.), *Psycholinguistic papers*. Edinburgh: Edinburgh University Press.

Krashen, S. & Scarcella, R. (1978). On routines and patterns in language acquisition and performance. *Language Learning, 28,* 283–300.

Krämer, I. (2005, May). *Egocentrism, deixis or conflict of domains? Lack of anaphoricity in child language*. Paper presented at the Workshop on Language Acquisition Between Sentence and Discourse, University of Nijmegen.

Kupisch, T. (2007). Testing the effects of frequency on the rate of learning: Determiner use in early French, German and Italian. In I. Gülzow & N.

Gagarina (Eds.), *Proceedings of the Workshop on Input Frequencies in Acquisition* (pp. 83–113). Mouton de Gruyter: SOLA Series.

Kuribara, C. (1999). Resettability of a syntactic parameter: Acquisition of category D in English by adult Japanese speakers. In C. Purdue (Ed.), *Proceedings of 8th EUROSLA Conference: Vol. 2* (pp. 119–132). Paris: British Institute in Paris.

Lakshmanan, U. (1993/1994). '"The boy for the cookie"—Some evidence for the nonviolation of the Case Filter in child second language acquisition'. *Language Acquisition, 3*, 55-91.

Lakshmanan, U. (1995). Child second language acquisition of syntax. *Studies in Second Language Acquisition, 17,* 301–329.

Langacker, R. (1987). *Foundations of Cognitive Grammar: Vol. 1*. Stanford: Stanford University Press.

Langacker, R. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 127–161). Philadelphia: John Benjamins.

Lardiere, D. (2000). Mapping features to form in second language acquisition. In J. Archibald (Ed.) *Second Language Acquisition and Linguistic Theory* (pp. 102–129). Malden, MA: Blackwell.

Lardiere, D. (2004). Knowledge of definiteness despite variable article omission in second language acquisition. In A. Brugos, L. Micciulla & C. Smith (Eds.), *Proceedings of the 28th Boston University Conference on Language Development* (pp. 328–339). Somerville, MA: Cascadilla Press.

Lardiere, D. (2005). On morphological competence. In L. Dekydtspotter, R. Sprouse & A. Liljestrand (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference* (pp. 178-192). Somerville, MA: Cascadilla Proceedings Project.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research, 25,* 173–227.

Leung, Y.-K. I., (2003). Failed Features versus Full Transfer Full Access in the Acquisition of a Third Language: Evidence from Tense and Agreement. In J. M. Liceras, H. Zobl & H. Goodluck (Eds.), *Proceedings of the 6th*

Generative Approaches to Second Language Acquisition Conference (pp. 199–207). Somerville, MA: Cascadilla Proceedings Project.

Leung, Y.-K. I. (2005). L2 vs. L3 initial state: A comparative study of the acquisition of French DPs by Vietnamese monolinguals and Cantonese-English bilinguals. *Bilingualism: Language and Cognition, 8,* 39–61.

Li, C. & Thompson, S. (1997). *Mandarin Chinese. A functional reference grammar*. Taipei, Taiwan: Crane Publishing.

Li, Y.-H. A. (1999). Plurality in a classifier language. *Journal of East Asian Linguistics, 8*, 75–99.

Lieven, E. (2009). Developing constructions. *Cognitive Linguistics, 20,* 191-199.

Lieven, E. (2008). Learning the English auxiliary: A usage-based approach. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. 60–98). Amsterdam: John Benjamins.

Lieven, E. & Tomasello, M. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. Ellis (Eds), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 168–196). New York: Routledge.

Lightfoot, D. (1991). *How to set parameters: Arguments from language change.* Cambridge, MA: MIT Press.

Lin, H. (2001). *A grammar of Mandarin Chinese*. Lincom Europa.

Lu, C. F.-C. (2001). The acquisition of English articles by Chinese learners. *Second Language Studies, 20,* 43–78.

Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.

Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21,* 557–587.

MacLaughlin, B. (1978). *Second language acquisition in childhood*. Hillsdale, NJ: Erlbaum.

MacWhinney, B. (1999). *Emergentist perspectives on language acquisition*. Mahwah, NJ: Erlbaum.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk (3rd edition)*. Mahwah, NJ: Erlbaum.

Mallinson, G. (1986). *Rumanian*. London: Croom Helm.

Marinis, T. (2003). *The Acquisition of the DP in Modern Greek*. Amsterdam: John Benjamins.

Matthews, D., Lieven, E., Theakston, A. & Tomasello, M. (2005). The role of frequency in the acquisition of English word order. *Cognitive Development, 20,* 121–136.

Matthews, S. & Yip, V. (1994). *Cantonese: A comprehensive grammar*. New York: Routledge.

Meisel, J. (2008). Child second language acquisition or successive first language acquisition? In B. Haznedar & E. Gavruseva (Eds.), *Current trends in child second language acquisition: A generative perspective* (pp. 55–80) Amsterdam: John Benjamins.

Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*. Amsterdam: John Benjamins.

O'Grady, W. (2008). Language without grammar. In N. Ellis & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 139–168). New York and London: Routledge.

O'Grady, W., Lee, M., & Kwak, H.-Y. (2009). In W. Ritchie & T. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 69–86). Bingley, UK: Emerald Press.

Paradis, J. (2005). Grammatical morphology in children learning English as a second language: Implications of similarities with Specific Language Impairment. *Language, Speech and Hearing Services in the Schools, 36,* 172–187.

Paradis, J. (2007). Second language acquisition in childhood. In E. Hoff & M. Shatz (Eds.), *Handbook of language development* (pp. 387–406). Oxford: Blackwell.

Paradis, J., Rice, M., Crago, M., & Marquis, J. (2008). The acquisition of tense in English: Distinguishing child L2 from L1 and SLI. *Applied Psycholinguistics, 29,* 1–34.

Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning, 37,* 361–83.

Pine, J. M., & Lieven, E. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics, 18,* 123–138.

Pine, J., Lieven, E. & Rowland, C. (1998). Comparing different models of the development of the English verb category. *Linguistics, 36,* 807–830.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1999). *Words and rules*. New York: Harper Collins.

Pizzuto E. & Caselli, M. C. (1992). The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language, 19,* 491–557.

Ravem, R. (1978). Two Norwegian children's acquisition of English syntax. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 148–154). Rowley, MA: Newbury House Publishers.

Rice, M. & Wexler, K. (2001). *Test of early grammatical impairment*. New York: The Psychological Corporation.

Rice, M., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language and Hearing Research, 41,* 1412–1431.

Richards, B. (1990). *Language development and individual differences: A study of auxiliary verb learning*. Cambridge: Cambridge University Press.

Robertson, D. (2000). Variability of the use of the English article system by Chinese learners of English. *Second Language Research, 16,* 135–172.

Robinson, P. & Ellis, N. (2008) (Eds.) *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge.

Rowland, C. (2007). Explaining errors in children's questions. *Cognition, 104,* 106–134.

Rowland, C. & Pine, J. (2000). Subject-auxiliary inversion errors and wh-question acquisition: "What children do know?". *Journal of Child Language, 27,* 157–181.

Rowland, C., Pine, J., Lieven, E. & Theakston, A. (2005). The Incidence of Error in Young Children's Wh-Questions. *Journal of Speech, Language, and Hearing Research, 48,* 384–404.

Ryding, K. C. (2006). *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Santelmann, L., Berk, S., Austin, J., Somarshekar, S., & Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no questions: Dissociating movement and inflection. *Journal of Child Language, 29,* 813–842.

Sarko, G. (2008). Accounting for indefinite article a drop in the oral L2 English of L1 Syrian Arab speakers. *Essex Graduate Student Papers in Language and Linguistics, 10,* 97–115.

Schmidt, R. L. (1999). *Urdu: An essential grammar*. Routledge.

Schneider, P. (2004a). Introducing the Edmonton Narrative Norms Instrument. Presentation at the Canadian Speech-Language Pathologists and Audiologists Conference, May 2004. Retrieved from http://www.rehabmed.ualberta.ca/spa/enni/.

Schneider, P. (2004b). The Edmonton Narrative Norms Project: Use of referring expressions. Poster presented at the Symposium for Research in Child Language Disorders, Madison, WI, June 2004. Retrieved from http://www.rehabmed.ualberta.ca/spa/enni/.

Schwartz, B. & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition: A longitudinal study of L1 Turkish German interlanguage. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 317–368). Amsterdam: John Benjamins.

Schwartz, B. & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12,* 40–72.

Schwartz, B. (2003). Child L2 acquisition: Paving the way. In B. Beachley, A. Brown & F. Conlin (Eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development* (pp. 26–50). Somerville, MA: Cascadilla Press.

Schwartz, B. (2004). Why child L2 acquisition? In J. Van Kampen & S. Baauw (Eds.). *Proceedings of Generative Approaches to Language Acquisition Conference 2003* (pp. 47–66). Utrecht, The Netherlands: LOT.

Selinker, L. & Lakshmanan, U. (1992). Language transfer and fossilization: The 'Multiple Effects Principle'. In S. Gass & L. Selinker (Eds.), Language Transfer in *Language Learning* (pp. 197–216). Amsterdam: John Benjamins.

Siebenborn, A.-K. (2010, June). *Schematization and generalization in novel construction learning (L1)*. Paper presented at the Child Language Seminar, City University of London.

Slabakova, R. (2000). L1 transfer revisited: The acquisition of telicity marking in English by Spanish and Bulgarian native speakers. *Linguistics, 38,* 739–70.

Snape, N, Leung, Y.-K. I. & Ting, H.-C. (2006). Comparing Chinese, Japanese and Spanish Speakers in L2 English Article Acquisition: Evidence against the Fluctuation Hypothesis? In M. G. O'Brien, C. Shea & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference* (pp. 132–139). Somerville, MA: Cascadilla Proceedings Project.

Tabors, P. O. (2007). *One child, two languages: A guide for preschool educators of children learning English as a second language*. Baltimore: Paul H. Brookes Publishing.

Theakston, A. & Lieven, E. (2008). The influence of discourse context on children's provision of auxiliary BE. *Journal of Child Language, 35*, 129–158.

Theakston, A. & Rowland, C. (2009a). The acquisition of auxiliary syntax: A longitudinal elicitation study. Part 1: Auxiliary BE. *Journal of Speech, Language, and Hearing Research, 52,* 1449–1470.

Theakston, A. & Rowland, C. (2009b). The acquisition of auxiliary syntax: A longitudinal elicitation study. Part 2: The modals and auxiliary DO. *Journal of Speech, Language, and Hearing Research, 52,* 1471–1492.

Theakston, A., Lieven, E., Pine, J. & Rowland, C. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics, 16*, 247–277.

Theakston, A., Lieven E. & Tomasello, M. (2003). The role of input in the acquisition of third-person singular verbs in English. *Journal of Speech, Language, and Hearing Research, 46*, 863-877.

Thomas, M. (1989). The acquisition of English articles by first- and second-language learners of English. *Applied Psycholinguistics, 10,* 335–355.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4,* 156–163.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. & Brooks, P. (1999). Early syntactic development. In M. Barrett (Ed.), *The development of language.* Hove, UK: Psychology Press.

Trenkic, D. (2007). Variability in L2 article production: beyond the representational deficit vs. processing constraints debate. *Second Language Research, 23,* 289-327.

Unsworth, S. (2005). *Child L2, Adult L2, Child L1: Differences and Similarities. A study on the acquisition of direct object scrambling in Dutch*. Doctoral dissertation, Utrecht University, The Netherlands.

Vainikka, A. & Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research, 12,* 7–39.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition, 40,* 21–81.

Valian, V. (2009). Abstract linguistic representations and innateness: The development of determiners. In W. Lewis (Ed.), *Time and again: Theoretical perspectives on formal linguistics. In honor of D. Terence Langendoen* (pp. 189–209). John Benjamins.

Valian, V., S. Solt & J. Stewart (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language, 36,* 743–778.

Wakabayashi, S. (2009). Lexical learning in second language acquisition: Optionality in the numeration. *Second Language Research, 25,* 335–341.

Wexler, K. (1998). Very early parameter setting and the Unique Checking Constraint: A new explanation of the Optional Infinitive Stage. *Lingua, 106,* 23-79.

White, L. (2003). Fossilization in steady state L2 grammars: Implications of persistent problems with inflectional morphology. *Bilingualism: Language and Cognition, 6,* 128–141.

White, L. (2007). Different? Yes. Fundamentally? No. Definiteness Effects in the L2 English of Mandarin Speakers. In R. Slabakova, J. Rothman, P. Kempchinsky, & E. Gavruseva. (Eds.) *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference 2007* (pp. 251–261). Somerville, MA: Cascadilla Proceedings Project.

Whong-Barr, M. & Schwartz, B. (2002). Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition, 24,* 579–616.

Wilson, S. (2003). Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language, 30,* 75–115.

Wong Fillmore, L. (1979). Individual differences in second language acquisition. In C. Fillmore, D. Kempler, & W.S.-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 203-227). San Diego, CA: Academic Press.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Yang, G. (1995). Aspect licensing, verb movement and feature checking. *Cahiers de Linguistique: Asie Orientale, 24,* 49–83.

Yuan, B. (1998). Interpretation of binding and orientation of the Chinese reflexive *ziji* by English and Japanese speakers. *Second Language Research, 14,* 324–40.

Zagona, K. (2002). *The syntax of Spanish*. Cambridge University Press.

APPENDICES

Appendix 1
In the longitudinal study, individual accuracy scores for the indefinite article (Figures 5.1a and 5.1b)

	Percent correct A					Percent correct THE				
[+article]	**R1**	**R2**	**R3**	**R4**	**R5**	**R1**	**R2**	**R3**	**R4**	**R5**
CNDX	83%	86%	89%	67%	60%	44%	78%	100%	100%	70%
DNNC	-	83%	71%	100%	75%	-	100%	100%	100%	97%
MRSS	0%	33%	50%	80%	50%	43%	96%	88%	100%	88%
JNNH	0%	-	20%	83%	50%	63%	-	100%	98%	100%
TNYN	20%	60%	60%	100%	100%	80%	100%	100%	93%	100%
DNNS	100%	78%	89%	50%	94%	60%	100%	100%	95%	100%
DNLN	60%	100%	25%	40%	50%	100%	100%	97%	96%	96%
RNL	0%	100%	67%	67%	58%	50%	83%	100%	89%	95%
GSYN	50%	60%	67%	86%	88%	43%	45%	74%	84%	100%
RMLM	0%	67%	75%	50%	82%	59%	94%	100%	83%	100%
mean	**35%**	**74%**	**61%**	**72%**	**71%**	**60%**	**88%**	**96%**	**94%**	**95%**
[-article]										
DVDC	20%	50%	100%	100%	100%	93%	100%	100%	100%	96%
SBST	-	78%	67%	83%	94%	0%	100%	91%	100%	100%
SMNS	75%	50%	55%	100%	90%	100%	100%	97%	100%	100%
LLKC	0%	0%	20%	89%	50%	100%	100%	100%	100%	100%
TRRK	50%	25%	0%	100%	88%	0%	88%	90%	100%	96%
YSSF	43%	100%	56%	83%	89%	67%	100%	88%	100%	100%
CHRS	57%	89%	100%	100%	100%	93%	100%	98%	100%	95%
mean	**41%**	**56%**	**57%**	**94%**	**87%**	**65%**	**98%**	**95%**	**100%**	**98%**

Appendix 2
In the longitudinal study, tokens and percentages of article errors out of the total number of contexts (Figures 5.2a and 5.2b)

[-article]		**R1**		**R2**		**R3**		**R4**		**R5**	
Indefinite contexts	*THE*	13/32	41%	12/61	20%	18/60	30%	14/48	29%	28/119	24%
	null	5/32	16%	4/61	7%	2/60	3%	2/48	4%	3/119	3%
Definite contexts	*A*	6/163	4%	0/178	0%	5/251	2%	4/222	2%	0/440	0%
	Null	56/163	34%	18/178	10%	9/251	4%	6/222	3%	27/440	6%
[+article]		**R1**		**R2**		**R3**		**R4**		**R5**	
Indefinite contexts	*THE*	17/32	53%	11/33	33%	14/48	31%	3/34	9%	10/88	11%
	null	0/32	0%	0/22	0%	2/48	4%	0/34	0%	0/88	0%
Definite contexts	*A*	3/66	5%	0/178	0%	0/125	0%	0/106	0%	1/163	1%
	null	2/66	3%	1/44	2%	6/125	5%	0/106	0%	2/163	1%

Appendix 3

In the cross-sectional study, individual scores for correct and incorrect article use in definite and indefinite contexts (Figures 5.3 and 5.4). Participants' name codes are grouped by L1 (Arabic, Mandarin/Cantonese, Spanish, and Hindi/Urdu/Punjabi

	Indefinite contexts							**Definite contexts**						
Name	***A***		***THE***		**Null**		**Total**	***THE***		***A***		**Null**		**Total**
008ma	1	13%	4	50%	3	38%	8	25	60%	0	0%	17	40%	42
024ma	9	100%	0	0%	0	0%	9	9	18%	0	0%	40	80%	50
036ma	5	71%	0	0%	2	29%	7	26	28%	0	0%	66	72%	92
043ma	1	14%	0	0%	6	86%	7	13	32%	3	7%	25	61%	41
044ma	7	70%	2	20%	1	10%	10	22	92%	1	4%	1	4%	24
058ma	6	86%	1	14%	0	0%	7	5	83%	0	0%	1	17%	6
072ca	1	13%	7	88%	0	0%	8	32	94%	1	3%	1	3%	34
081ma	3	27%	7	64%	1	9%	11	62	98%	0	0%	1	2%	63
099ma	1	.	1	.	0	.	2	54	89%	0	0%	7	11%	61
138ca	11	79%	2	14%	1	7%	14	40	98%	1	2%	0	0%	41
mean		52%		28%		18%			69%		2%		29%	
040ur	1	13%	7	88%	0	0%	8	38	90%	0	0%	4	10%	42
042ur	7	54%	6	46%	0	0%	13	57	100%	0	0%	2	4%	57
061ur	9	82%	2	18%	0	0%	11	15	94%	0	0%	1	6%	16
071ur	5	50%	1	10%	4	40%	10	0	0%	2	4%	42	91%	46
082pu	4	36%	1	9%	6	55%	11	29	59%	3	6%	17	35%	49
088pu	12	60%	8	40%	0	0%	20	65	78%	0	0%	18	22%	83
134hi	4	50%	2	25%	2	25%	8	3	17%	0	0%	15	83%	18
155pu	0	0%	2	50%	2	50%	4	1	11%	0	0%	8	89%	9
162ur	8	57%	6	43%	0	0%	14	9	82%	1	9%	1	9%	11
163ur	18	95%	1	5%	0	0%	19	16	89%	1	6%	1	6%	18
mean		50%		33%		17%			62%		3%		35%	
004sp	2	22%	7	78%	0	0%	9	10	100%	0	0%	0	0%	10
019sp	17	94%	1	6%	0	0%	18	41	91%	2	4%	2	4%	45
020sp	11	85%	2	15%	0	0%	13	19	70%	0	0%	8	30%	27
105sp	2	33%	3	50%	1	17%	6	13	93%	0	0%	1	7%	14
106sp	1	33%	2	67%	0	0%	3	26	96%	1	4%	0	0%	27
111sp	8	62%	4	31%	1	8%	13	29	100%	0	0%	0	0%	29
128sp	4	57%	3	43%	0	0%	7	16	76%	1	5%	4	19%	21
139sp	8	53%	7	47%	0	0%	15	16	100%	0	0%	0	0%	16
141sp	10	71%	4	29%	0	0%	14	25	100%	0	0%	0	0%	25
164sp	1	14%	6	86%	0	0%	7	42	98%	0	0%	1	2%	43
mean		53%		45%		2%			92%		1%		6%	
028ar	7	58%	5	42%	0	0%	12	27	100%	0	0%	0	0%	27
051ar	1	14%	5	71%	1	14%	7	24	92%	0	0%	2	8%	26
052ar	5	45%	6	55%	0	0%	11	19	100%	0	0%	0	0%	19
056ar	12	75%	2	13%	2	13%	16	34	100%	0	0%	0	0%	34
065ar	1	20%	4	80%	0	0%	5	13	93%	0	0%	1	7%	14
066ar	4	44%	5	56%	0	0%	9	40	100%	0	0%	0	0%	40
075ar	2	40%	2	40%	1	20%	5	7	100%	0	0%	0	0%	7
093ar	8	89%	1	11%	0	0%	9	11	100%	0	0%	0	0%	11
135ar	3	25%	8	67%	1	8%	12	33	89%	2	5%	2	5%	37
136ar	11	79%	3	21%	0	0%	14	27	100%	0	0%	0	0%	27
mean		49%		45%		6%			97%		1%		2%	

Appendix 4

In the cross-sectional study (TEGI), individual accuracy scores for BE and DO. Uninverted questions were excluded from the analysis

Declaratives with BE				**Questions with BE**			**Questions with DO**		
Name	corr	total	% corr	corr	total	% corr	corr	total	% corr
066AR	4	6	67%	2	5	40%	2	10	20%
056AR	7	7	100%	4	6	67%	10	12	83%
075AR	7	7	100%	5	6	83%	4	6	67%
095AR	6	6	100%	1	4	25%	0	2	.
091AR	7	7	100%	0	0	.	0	1	.
093AR	6	6	100%	5	6	83%	6	11	55%
052AR	6	6	100%	2	6	33%	2	11	18%
096AR	9	10	90%	3	5	60%	1	5	20%
136AR	5	5	100%	3	4	75%	1	12	8%
028AR	6	6	100%	0	4	0%	1	10	10%
097AR	1	4	25%	0	4	0%	0	1	.
135AR	7	7	100%	4	6	67%	6	11	55%
mean			**90%**			**48%**			**37%**
036MA	2	6	33%	2	6	33%	0	11	0%
058MA	7	7	100%	4	6	67%	0	10	0%
044MA	6	7	86%	0	5	0%			.
024MA	0	5	0%	0	2	0%			.
016MA	4	9	44%	4	6	67%	0	11	0%
138CH	2	4	50%	5	5	100%	3	11	27%
026CA	1	2	.	0	6	0%			.
081CH	2	5	40%	1	6	17%	0	11	0%
090CH	3	5	60%	3	5	60%	0	9	0%
048MA	7	7	100%	3	6	50%	3	10	30%
143CH	6	6	100%	4	5	80%	1	11	9%
076CH	5	6	83%	2	5	40%	4	11	26%
mean			**62%**			43%			**10%**
004SP	2	2	.	0	2	.			-
005SP	1	2	.	0	1	.			-
019SP	7	7	100%	0	1	.			-
020SP	8	8	100%	5	5	100%	5	11	45%
105SP	6	6	100%	3	6	50%	4	10	40%
106SP	5	5	100%	3	3	100%	9	9	100%
107SP	3	5	60%	1	2	.	1	2	.
128SP	1	4	25%	0	4	0%	1	2	50%
111SP	6	7	86%	6	6	100%	11	11	100%
139SP	0	0	.	6	6	100%	4	9	44%
141SP	7	7	100%	5	6	83%	7	11	64%
167SP	7	7	100%	1	6	17%	6	12	50%
mean			**84%**			**55%**			**60%**

Appendix 4 (continued)

	Declaratives with BE			**Questions with BE**			**Questions with DO**		
	corr	total	% corr	corr	total	% corr	corr	total	% corr
042UR	8	8	100%	1	4	25%	2	10	20%
037UR	8	8	100%	1	4	25%	5	10	50%
027PU	4	5	80%	0	2	.	0	2	0%
061UR	4	5	80%	0	2	.	7	11	64%
071UR	4	5	80%	0	3	0%	0	2	.
088PU	5	5	100%	3	4	75%	2	10	20%
089UR	2	3	67%	0	5	0%	0	2	.
150PU	1	3	33%	0	3	0%	.	.	.
161UR	3	3	100%	0	3	0%	1	1	.
163UR	4	5	80%	4	8	50%	1	8	13%
176UR	6	6	100%	0	1	.	0	2	.
169PU	7	7	100%	1	6	17%	6	12	50%
mean			**85%**			**16%**			**29%**