Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.

2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.

3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

INTERFACE IN THE ACQUISITION OF LANGUAGE FORM

AND CONTENT BY MANDARIN-SPEAKING CHILDREN

TANG PO YI

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Interface in the Acquisition of Language Form and Content

by Mandarin-Speaking Children

TANG Po Yi

A thesis submitted in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

August 2023

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

_____TANG Po Yi_____ (Name of student)

Abstract

Children with language disorders experience significant functional impairment in their daily lives, underscoring the importance of appropriate intervention to minimize the negative impacts. Their language disruptions can manifest across different language domains, resulting in a heterogeneity among children with language disorders. Thoroughly examining various language domains is crucial to comprehensively understand their language profiles, such that their diverse needs can be addressed accordingly. This thesis adopts the content-form framework to investigate the language profiles of Mandarin-speaking children, both with and without language disorders, focusing on the domains of content and form. It also aims to gain insights into potential disruptions in expressing semantic content categories, as well as the content-form interface. By utilizing this framework, traditional language measures are supplemented, providing comprehensive language profiling and guiding intervention directions for children with language disorders.

To document the language productions in the domains of both content and form in Mandarin-speaking children, the Corpus of Mandarin Child Language (CMCL) was established in study one. Naturalistic language samples were collected from 82 native Mandarin-speaking children aged 25 to 60 months. Semantic content categories expressed in each utterance were tagged, along with the annotations on parts-of-speech, to investigate the acquisition trajectory of various semantic content categories. The strong correlations between age and traditional language measures such as mean length of utterance and lexical diversity, are aligned with the literature, validating the CMCL. Additionally, the acquisition trajectory of semantic content categories and the way it is influenced by factors of syntactic and cognitive complexity were examined. As a pilot test, the interaction between content and form in the acquisition process was also explored by observing children's expressions of different semantic

content categories using specific form (verb), and different forms expressing a particular semantic content category (*dative*) from the CMCL.

In study two, a particular semantic content category - *temporal* - was selected for more detailed investigation on the content-form interface in child language acquisition. To explore how Mandarin-speaking children acquire different forms to express the same content, the acquisition between three different temporal-marker groups, namely aspect markers (AMs), temporal adverbs (TAs) and temporal nouns (TNs), by Mandarin-speaking children were examined. Results indicated that various time concepts regulated the acquisition trend among the three different temporal markers, influencing the semantic representations of time and the acquisition of different temporal-marker groups. Next, to examine how Mandarin-speaking children's acquisition of lexical forms is affected by content, the acquisition within each temporal-marker group was also examined. Results indicated that the acquisition of different lexical items within each group also appeared to be affected by the semantic features of temporal remoteness and specificity, with more remote and specific items acquired later. These findings highlight the interaction of content and form in modulating the acquisition of temporal markers in Mandarin.

To illustrate the content-form interface idea in language disorder, case studies were conducted on two Mandarin-speaking children with language difficulties (LD) in study three, comparing their language samples with those of typically developing children (TD) from study one. LD children demonstrated shorter mean length of utterance and lower lexical diversity than their TD peers, confirming their language difficulties. Disruptions in expressing certain semantic content categories, which might manifest through the content-form interface, were also observed. Examining semantic content category and its content-form interface thus enriches the language profile of children with language disorders, and allows the identification of subtle disruptions that may not be easily captured through traditional measures alone.

Overall, these findings enhance our understanding of the language profiles of Mandarin-speaking children with and without language disorders, particularly from a semantic perspective, as well as the content-form interface. Empirically, they contribute experimental data on the acquisition of semantic content categories in Mandarin-speaking children. The CMCL also provides a platform for examining the content domains and the content-form interface in early child language acquisition. Theoretically, the bidirectional interaction between language-general content and the language-specific form is suggested, with the content-form interface regulating the language acquisition of young children. Clinically, intervention direction for children with language disorders is informed by a thorough consideration of the content and form domains, as well as the content-form interface.

Publications

Tang, T. P. Y., Hsu, Y. Y., Lau, D. K. Y., & Leung, M. T. (2024). Content-form interaction
in the acquisition of temporal markers by Mandarin-speaking children. *SAGE*
Open, *14*(1). https://doi.org/10.1177/21582440241234215

Tang, T. P. Y., Lau, D. K. Y., & Leung, M. T. (2023). Corpus of Mandarin Child Language:
a preliminary study on the acquisition of semantic content categories in Mandarin-
speaking preschoolers. *Frontiers in Psychology*, *14*, 1234525. doi:
10.3389/fpsyg.2023.1234525

Acknowledgments

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviations	Items
AM	Aspect Marker
CHILDES	Child Language Data Exchange System
CLAN	Computerized Language Analysis
CMCL	Corpus of Mandarin Child Language
ET	Event Time
LD	Language Difficulty
LSA	Language Sample Analysis
MLU	Mean Length of Utterance
NDW	Number of Different Words
RT	Reference Time
SES	Socioeconomic Status
SFP	Sentence Final Particle
ST	Speech Time
TA	Temporal Adverb
TD	Typical development
TN	Temporal Noun
TNU	Total Number of Utterances
TNW	Total Number of Words
TTR	Type-Token Ratio

Chapter One

Introduction

Communication is an essential part of human life. We use language as a unique tool to communicate effectively. Language involves the use of arbitrary signals (e.g., words, signs) in a systematic way, allowing us to share ideas, information, feelings and thoughts about the world around us. From the moment of birth, children begin to explore the environment using their sensory and motor systems, and they rapidly acquire language from their social environment (Adolph & Berger, 2006). As children grow, they gradually develop the ability to use words, phrases and more complex sentences during the preschool years. This enables them to express their needs, describe events in various daily contexts, and acquire new concepts and knowledge from the environment (Vygotsky, 1978). It also allows them to engage in social interactions and develop relationships with the people surrounding them. In view of the above, it is crucially important for children to acquire language so that various developmental milestones and daily functioning can be achieved.

However, despite relatively intact cognitive, neurological, motor and sensory abilities, a significant portion of children, estimated to be around 7-9%, experience unexpected difficulties in producing or understanding language (Bishop et al., 2017; Leonard, 2014). These language disruptions can endure over time and cause functional impairment in daily life. Without proper intervention, these difficulties may persist into adulthood, causing long-term adverse impacts on individuals with language disorders (Bishop et al., 2017). Understanding the specific language difficulties experienced by these individuals is crucial for planning and implementing appropriate interventions that can minimize the negative impacts they face. By gaining insights into the nature and extent of their language impairments, professionals can develop tailored intervention strategies that target the specific areas of need.

According to Olswang & Bain (1991), there are three main purposes of language intervention. The first objective of intervention is to change or eliminate the underlying problems contributing to language difficulties. Secondly, language intervention aims to improve discrete aspects of language function. Lastly, intervention also involves teaching compensatory strategies to individuals with language disorders. Expanding on these purposes, Paul et al. (2018) introduced a fourth objective of language intervention. This objective involves modifying the context to optimize the child's communicative potential. The above purposes collectively aim to support individuals with language disorders in developing their language skills and enhancing their overall communication abilities.

To accomplish the above objectives, language intervention also involves the selection of an intervention approach, treatment context and treatment goals (McCauley et al., 2016; Paul et al., 2018). Intervention approaches encompass a range of methods, strategies and techniques that are selected based on an underlying theoretical framework to support individuals with language difficulties. For example, the child-centered approach emphasizes learning language in natural, meaningful contexts through play and everyday activities (Paul et al., 2018). This approach focuses on following the child's interests and creating opportunities for language learning within the child's natural environment (Fey, 1986). Clinician-directed intervention, on the other hand, refers to a therapeutic approach where clinicians actively guide and facilitate a child's language training through structured and planned activities that specifically target the acquisition of language skills in children (Paul et al., 2018). Likewise, treatment context represents the physical and social environment in which language interventions take place. The physical context encompasses physical space like the child's home, and any visual supports and assistive technology present. Additionally, the social context is shaped by communication partners and the linguistic and cultural factors at play.

The third component of the language intervention – goal – represents the linguistic knowledge and communicative behaviors that are expected to be achieved. It guides intervention on how to make changes in the communication abilities of children with language disorders (Fey, 1986). According to McCauley et al. (2016), there are three different types of treatment goals. Basic goals identify the area of communication which is functionally important. Intermediate goals specify the area within each basic goal while specific goals further target specific examplars of the language aspects. Since each language has its unique characteristics, structures and rules, treatment goals in the intervention for children with language disorders should therefore be language-specific. By focusing on the specific language difficulties that a child is experiencing within different aspects of language, interventions can be tailored to address their specific needs effectively and ensure effective progress. Besides, children with language disorders exhibit a certain degree of heterogeneity, meaning that their language difficulties and responses to various language interventions can vary significantly from one child to another (Law et al., 2003). This heterogeneity poses a challenge in developing and implementing effective language interventions that meet the diverse needs of these children. Given the above, it is important to have a thorough examination of various language domains to gain a comprehensive understanding of the language profiles of children with langauge disorders. This allows professionals to identify their unique strengths and weaknesses within different language domains, facilitating the development of specific intervention goals. Additionally, compensation strategies may also be considered to support effective communication.

In order to reduce the adverse effects of language disorder in children, a thorough understanding of their specific language difficulties across different domains is crucial. The proposal of Content, Form and Use by Bloom and Lahey (1978) may provide such a framework to profile their language abilities for the identification of relative strengths and weaknesses.

Therefore, this thesis adopts the content-form framework and compares the language profiles of Mandarin-speaking children, with or without language difficulties. The domains of content and form, as well as the content-form interface[1], were investigated in order to identify the relative strengths and weaknesses in children with language disorder. In this introductory chapter, the three language domains – content, form and use (Bloom & Lahey, 1978) – are reviewed in order to gain a more comprehensive understanding of their roles played in child language. After that, literature on acquisition studies examining the domains of content and form in child language is also examined. The importance of content-form interaction in child language studies is also highlighted. Following these, different experimental methods in studying child language are also evaluated. Finally, the motivations and objectives of the study are presented, while the outline of this thesis can also be found towards the end.

1.1 Content, Form and Use of Language

Language can be defined as a standardized set of symbols and the knowledge about how to combine those symbols into words, sentences, and texts to convey ideas and feelings (Gillam & Marquardt, 2016). To better understand child language disorder, the linguistic knowledge of different domains of language should be considered. According to Bloom and Lahey (1978), Content, Form and Use are three major domains of language. Language content refers to the meaning and information conveyed through language, encompassing ideas and concepts about the world expressed in spoken or written forms, such as words, phrases, sentences and discourse. It encompasses notions about objects, actions, and relations between objects and events, allowing us to express virtually everything that we know about the world

---

[1] The interface suggested here refers to the working space for integrating the knowledge of semantic meaning and linguistic form together. This interface enables bidirectional processing of information so that children can use different forms to express a particular content, or vice versa.

(Gillam & Marquardt, 2016). While language content reflects different topics of messages, it remains broad and applicable across different languages and contexts. However, particular language topics may vary with age and culture (Bloom & Lahey, 1978).

The second domain, known as language form, pertains to the surface structure of language. It represents the systems of rules that govern the arrangement of particular language features and the combinations of different linguistic units within messages. Language form comprises three broad categories: phonology, syntax and morphology, which collectively contribute to the underlying rule system of language (Bloom & Lahey, 1978; Gillam & Marquardt, 2016). Phonology represents the sounds of spoken language and encompasses rules governing the organization of segmental and suprasegmental features in speech. Morphology concerns the rules governing the formation of words through the combination of meaningful units known as morphemes (Matthews, 1974). Syntax, on the other hand, represents the rules governing the arrangement of words within sentences (Barrett, 1999).

Language use refers to the goals for communication, as well as how people choose between alternative combinations of words and sentences to convey those objectives within various contexts (Gillam & Marquardt, 2016). According to Bloom & Lahey (1978), there are three key components of language use. Firstly, the goals and functions of language provide the motivations and reasons for us to speak and listen to others. Secondly, consideration of contextual information allows us to achieve different intended language goals effectively. Finally, the interactions between persons also provide cues and guidelines on how to initiate, maintain and terminate conversations with communication partners.

The above three language domains provide a systematic framework to profile and study the language of children with or without language disorders. In this thesis, the domains of content and form, as well as their interface are the main foci of investigation, whereas the

domain of use is controlled in data collection, with different standardized elicitation procedures, and not manipulated in the experiment. Besides, as mentioned before, language involves the expressions with different symbols. Language contents can be expressed not only through spoken sounds but also through objects, pictures, words and other conventional signs etc. (Barrett, 1999). In this thesis, while language use is beyond the scope of the study, the terms "form" and "content" are used in a more restricted sense to denote the spoken words used by young children to encode the early ideas and topics about the world.

1.2 Language Acquisition in Young Children

To effectively plan intervention for children with language disorders, it is valuable to compare their language abilities with those of typically developing children. This allows the identification of strengths and weaknesses in children with language disorders and helps shape intervention goals. As such, language profiling enables a thorough examination of various language aspects in both children with or without language disorders. By conducting language profiling, practitioners can gain insights into different language domains and compare the language skills of children with language disorders to those of typically developing children. This comparison provides a basis for understanding the specific areas of difficulty and the unique needs of children with language disorders. To establish a comprehensive understanding of how language is acquired in early childhood, acquisition studies focusing on the domains of form and content in typically developing children are reviewed in the following sections.

1.2.1 Studies of Form in Child Language Acquisition

To begin, previous studies of child language have focused extensively on the acquisition of various categories of language forms. To examine the category of phonology, some studies have investigated the acquisition trajectory of various phonemes by young children (e.g., MacLeod et al., 2011; To et al., 2013). Besides, the acquisition of different suprasegmental features including intonation (e.g., Demuth, 2015; Esteve-Gibert & Prieto, 2014) and lexical tones (e.g., Li & Thompson, 1977; So & Dodd, 1995) were also examined. Likewise, considerable attention has also been devoted to studying the acquisition of morphology and syntax in particular. The emergence of different parts of speech in the early vocabulary of young children has been studied extensively (e.g., Bleses et al., 2008; Lieven et al., 1992). Besides, a sizable number of studies were also conducted to investigate the acquisition of various morphological and syntactic structures in young children. For instance, early studies on child language have established a consistent order of acquisition for different English inflections among young children (e.g., Brown, 1973; de Villiers & de Villiers, 1973). Likewise, Scarborough et al. (1985) examined various syntactic structures and found that children's syntactic complexity and diversity increased with age, indicating the acquisition of more advanced syntactic structures as children grew older. Similarly, Demuth & McCullough (2009) reported a gradual acquisition of more complex sentence structures with auxiliary verbs among preschool children as they matured.

Apart from the above, various coding schemes and indexes have also been developed for the analyses of various morphological and syntactic constructions. The Developmental Sentence Score (DSS; Lee & Canter, 1971), the Systematic Analysis of Language Transcripts (SALT; Miller & Chapman, 1985), the Index of Productive Syntax (IPSyn; Scarborough, 1990), and the Sampling Utterances and Grammatical Analysis Revised (SUGAR; Pavelko & Owens, 2017) are examples of language tools that provide systematic ways to assess and measure specific aspects of language form and structure in children.

Among studies in the acquisition of form, investigating the sentence length of children also constituted a crucial part of the literature. The longitudinal study on three children between the ages of 2 to 3 conducted by Brown (1973) is one of the most pioneer studies in sentence length. By analysing participants' language productions, Brown (1973) found that children's mean length of utterance in morphemes (MLU) increased with age. The children also progressed through stages of grammatical development with an increase in utterance length. Brown's findings were also replicated in other child language studies (e.g., Miller, 1981; Moyle et al., 2011; Paul, 2000; Rice et al., 2010). Although MLU is a useful measure of syntactic development, it was argued that it may not be sensitive enough to capture the full range of language abilities in children (e.g., Rice and Wexler, 1996; Rondal et al., 1987; Scarborough et al., 1991). Assessing sentence length alone seems to be insufficient for profiling the language abilities of children. While sentence length can provide some insights into a child's syntactic development, it is important to consider a wide range of language skills and factors. In order to understand child language thoroughly, a more comprehensive approach with multiple measures to study language acquisition is recommended.

1.2.2 Studies of Content in Child Language Acquisition

In addition to the study of form, the investigation of contents also plays a significant role in the field of child language acquisition. Bowerman (1973) conducted a study on her daughter's acquisition of spatial terms between the ages of 1;4 and 2;6. Her findings suggested that children's understanding of word meanings is closely tied to their perceptual and motor experiences with the world. Another study by Bloom (1970) observed children's natural language use and identified a set of early semantic relations commonly used by children to

convey meanings in their daily conversations. This study provides valuable insights into the significance of semantic knowledge in language acquisition.

In addition to earlier research, more recent studies on child language acquisition have also explored the content domain. Considerable research has been conducted to investigate the acquisition of early vocabulary in children. Jiménez et al. (2021) conducted a study investigating the lexical profiles of typically developing children using the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 1993). They analysed the production of twenty-two semantic categories to gain insights into the developmental milestones and patterns of children's vocabulary acquisition. Apart from the investigation of semantic categories, the impact of various semantic measures and features on language acquisition was also examined. Stockman and Vaughn-Cooke (1992) suggested that the acquisition of locative words was influenced by the meaning and relational roles of the words. Their study highlighted the importance of understanding the semantic properties and relational meanings of locative words in children's language acquisition, and provided insights into the developmental trajectory of spatial language skills in children. Similarly, Horvath et al. (2022) reported the impact of manner/result features of verbs on their acquisition in children. Accordingly, children often learn manner verbs earlier, as they tend to be more perceptually salient and observable. On the other hand, result verbs, which describe the outcome or change resulting from an action, are acquired later by children.

The above acquisition studies in semantic categories and various semantic measures of early vocabulary shed light on the significance of the content domain in the early stages of language acquisition, and offered an additional perspective to understand the language abilities of young children on top of the study of surface form. Nevertheless, while there has been significant research investigating the content domains in terms of semantic category and different semantic features of vocabulary, studies that go beyond the lexical-semantic level

have been relatively limited (e.g., Bloom, 1991; Green, 1975; Stockman & Vaughn-Cooke, 1986). (More details can be found in chapter two) As such, investigating semantic content beyond individual words may contribute to a more holistic understanding of children's language abilities. Furthermore, exploring content beyond the lexical level is also crucial for our understanding of language disorders. Researchers can identify the specific areas of difficulty and design targeted interventions to support children with language disorders.

1.2.3 Content-Form Interaction in Language Acquisition

Apart from the roles played by individual domains, Lahey and Bloom (1977) proposed that language acquisition also involves understanding the interplay between these three domains. They emphasized that children not only acquire linguistic forms but also develop an understanding of the content or meaning expressed by those forms and how to use language effectively in communication. Similarly, Gillam and Marquardt (2016) also highlighted that various aspects of language are interrelated. They emphasized that language form, including phonology, syntax, and morphology, serves as a means to connect sounds with meaning in spoken language. The combination of linguistic symbols (form) is influenced by the meaning relations (content) that children express and the purpose and context (use) of their utterances. By considering the interaction between content, form, and use, researchers can better understand how children develop their understanding of meaning, how they use language to communicate effectively, and how linguistic forms evolve to serve these purposes. In sum, it is important to examine the interaction between different domains in addition to the independent investigation of each language domain, so that a holistic understanding of child language can be attained.

Content-form interaction, in particular, refers to the complex interplay between semantic content and linguistic form in language acquisition. The form of language shapes the

interpretation of the content, whereas the content affects the selection of appropriate forms to convey the intended meaning. A notable example that demonstrates the content-form interface in early language acquisition is the bootstrapping strategies. One strategy that contributes to children's language acquisition is the concept of 'semantic bootstrapping' (Pinker, 1984). According to this approach, children utilize their knowledge of the meaning of words and the relationships between them to infer and acquire various grammatical rules and syntactic structures. They initially construct the semantic representations and then establish connections between these semantic entities to the syntactic properties of language forms. Likewise, the 'syntactic bootstrapping' approach (Gleitman, 1990) suggests that knowledge of syntax can facilitate the acquisition of word meanings. By analyzing the syntactic context in which a new word is used, children can make inferences about its meaning and incorporate it into their vocabulary. Both semantic and syntactic bootstrapping mechanisms are essential for early language acquisition. These mechanisms allow children to acquire language more effectively by actively engaging in the processing of connecting semantic and syntactic information. By leveraging their existing linguistic knowledge, children can make inferences and acquire language more effectively, ultimately contributing to their holistic understanding and acquisition of language. Considering the above, further exploration of the content-form interface will probably enhance our understanding of language acquisition, as well as language disorders in young children.

1.3 Experimental Methods in Studying Child Language

In the field of child language studies, researchers have utilized various experimental methods to examine the language abilities of children. Ambridge & Rowland (2013) suggested

that elicited imitation, syntactic priming and elicited production are among the most commonly used experimental methods within the production paradigm in child language studies.

Elicited imitation is a highly structured method used in child language studies. It involves children repeating sounds, words, phrases or utterances that are modeled by the examiner. Elicited imitation allows for customization of target language responses, as well as manipulating the difficulty level of the tasks. One major advantage of elicited imitation is that the linguistic input for children is controlled so that researchers can study specific aspects of language in a systematic manner (Gathercole & Baddeley, 1989).

Syntactic priming, another structured method commonly employed in language acquisition research, involves children producing sentences with a particular syntactic structure that corresponds to a preceding utterance modeled by the examiner. The syntactic priming task requires children to generate new utterances with the same structure but different content and lexicons. By processing the syntactic structure presented to them and subsequently formulating a new utterance with the same structure, children tend to develop a mental representation of the syntax involved (Pickering & Ferreira, 2008). Similar to elicited imitation, syntactic priming allows the elicitation of specific or less frequently encountered syntactic structures. This method provides a means to investigate children's syntactic learning processes.

The above two methods demonstrate a relatively structured approach to examining language, with a primary focus on the production of form. Alternatively, some approaches allow children to express themselves in less structured situations, such as elicited production. In elicited production, specially designed contexts are created, and the experimenter provides both verbal and nonverbal prompts to elicit children's responses. This method exists along a continuum, with one end involving productions elicited in a naturalistic context with the preservation of the communicative sense of production (Ambridge & Rowland, 2013). One

example of elicited production is language sample analysis (LSA), which involves analyzing spontaneous speech production from children. LSA is considered an ecologically valid and authentic method of assessment (Owens, 2010), and often displays a more naturalistic and representative picture of a child's language compared to standardized assessments (Evans & Craig, 1992). The administration for language sample taking is relatively simple, making it suitable for use with very young children. Additionally, LSA offers the advantage of examining children's language across different language domains, including semantics, syntax, morphology, and pragmatics (Overton & Wren, 2014). To this end, it is worth noting that many studies utilizing LSA predominately focus on form and lexical semantics. For instance, LSA has been employed to examine children's acquisition of vocabulary (e.g., Liu, 2007; Liu et al., 2008; Stoll et al., 2012), early grammatical constructions (e.g., Lee & Wong, 1998), syntactic complexity (e.g., Deng et al., 2018; Diessel, 2004; Lu, 2009), morphology (e.g., Jia & Fuse, 2007; Maslen et al., 2004), and discourse relationship (Zhou & Xue, 2015). In order to assess both content and form domains and provide a comprehensive and holistic understanding of children's language abilities, it is important for language sample analysis in eliciting productions to move beyond investigating surface form alone.

Language corpora, which consist of transcribed and usually annotated language samples, are also essential for the study of child language. Biber et al. (1998) stated that using corpora to study language is empirical as it involves analyzing a large collection of samples from real-life situations. The availability of computer analytical techniques further accelerates and enhances various quantitative and qualitative analyses of these language samples, covering different aspects of language. Quantitatively, researchers can calculate the distribution and properties of different structures in terms of types, tokens and percentages. Qualitatively, both target and non-target structures can also be investigated (Deng & Yip, 2018). In addition, corpora can be used multiple times and made publicly available for other researchers. Currently,

there are many cross-sectional language corpora available in various languages that employ LSA (e.g., Aguado-Orea & Pine, 2015; Li & Zhou, 2015; Tommerdahl & Kilpatrick, 2013). These corpora serve as valuable resources for studying child language acquisition. However, most of the existing corpora were primarily annotated syntactically, lacking annotations on semantic content. Notably, certain semantic content was not specifically elicited in the language samples of these databases, making it challenging to thoroughly examine language abilities from the semantic perspective.

As discussed earlier, the inclusion of semantic content in language studies adds valuable insights to the analysis of child language acquisition. While both content and form are crucial aspects of language, incorporating semantic annotations in language corpora offers an additional perspective for investigating child language. Sagae et al. (2010) have also suggested that the future direction of language studies should involve the incorporation of semantic information into the corpora. Therefore, it will be of great importance to establish a child language database that adopts a specially designed protocol for eliciting various semantic contents. This database would not only collect language samples, but provide corresponding semantic annotations, allowing for a comprehensive examination of child language in the semantic domain beyond the lexical level.

In light of the above, this study aims to establish a child language database annotated both syntactically with part of speech and semantically with semantic content category (Lahey, 1988). This database seeks to document child language in the semantic domain and provide a foundation for comparative analyses with children having language disorders, enabling researchers to gain insights into the differences and similarities between typically developing children and those with language disorders.

1.4 Motivations and Objectives

Mandarin, the official language of China, boasts the highest number of native speakers worldwide. Considering the prevalence of language disorders globally (Bishop et al., 2017; Leonard, 2014), a significant number of Mandarin-speaking children are also adversely affected. It is crucial to understand the language difficulties and to identify the strengths and weaknesses of Mandarin-speaking children with language disorders in order to develop effective intervention plans and set appropriate goals. To achieve this, it is necessary to profile language disruptions across different language domains and compare them to typically developing populations.

The above literature review highlighted the significance of examining different language domains and their interactions in the acquisition process to gain a better understanding of how children acquire language. While the importance of semantic contents in child language study has been acknowledged (e.g., Bowerman, 1973), much of the research in this area has primarily focused on investigating lexical semantics, specifically vocabulary and semantic features. The acquisition of the content domain beyond lexical level remains relatively unexplored. Besides, the content-form interface in language acquisition has received limited attention as well. Therefore, it would be worthwhile to supplement the study of language from the semantic perspective. This holistic approach would shed light on theoretical, empirical and clinical aspects related to language studies.

Given the above, the current study adopts the content-form framework (Bloom & Lahey, 1978) to examine the language abilities of Mandarin-speaking children, both with and without language disorders. The primary objective of this thesis is to imply directions for intervention for Mandarin-speaking children with language disorders. To accomplish this, the thesis encompasses several studies with different objectives. Firstly, a syntactically and semantically

annotated database adopting LSA is established to serve as a tool and platform for investigating the domains of content and form, as well as the content-form interface in child language. The acquisition of various semantic content categories (Lahey, 1988) by Mandarin-speaking children is examined within this framework. Besides, this thesis explores the significance of studying the content-form interface in the language acquisition of Mandarin-speaking children. It also enhances understanding of how semantic information interacts with the syntactic structures during language acquisition. Next, the thesis compares language production, considering content, form and the content-form interface between Mandarin-speaking children with typical development and two cases with language difficulties. This comparison provides insights into the strengths and weaknesses of children with language disorders in different language domains, aiming to provide a foundation for developing comprehensive interventions with appropriate treatment goals.

1.5 Overview of the Thesis

This thesis comprises three studies that investigate the acquisition of semantic content categories and the content-form interface in Mandarin-speaking children with or without language disorders. Several theoretical, empirical and clinical issues that have not been explored in previous literature on Mandarin child language acquisition and disorder are also addressed. The three studies are outlined as follows:

Study one establishes the Corpus of Mandarin Child Language (CMCL), which documents the production of different semantic content categories (Lahey, 1988) by Mandarin-speaking children. Naturalistic language samples were collected, and each utterance was annotated with semantic content categories and part-of-speech tags. The study examines the acquisition trajectory of various semantic content categories in typically developing Mandarin-

speaking children. Besides, traditional language measures, such as mean length of utterance and lexical diversity, are also analyzed. Additionally, the interaction between semantic content and language form is also investigated by observing the use of a specific form (verb) to express different semantic content categories, and the use of different forms to express a particular semantic content category (dative) (see Chapter two).

Study two focuses on a particular semantic content category - *temporal* - for a more detailed investigation of the content-form interface in child language acquisition. It investigates how Mandarin-speaking children acquire different forms to express the same content. The acquisition between three different temporal-marker groups, namely aspect markers (AMs), temporal adverbs (TAs) and temporal nouns (TNs), by Mandarin-speaking children is examined. In addition, the effects of temporal remoteness and specificity on the acquisition of different lexical terms within each category of temporal markers are also investigated. This study explores the acquisition patterns between and within each temporal-marker group and examines how the acquisition of forms is affected by content (see Chapter three).

Study three examines the manifestations of disruptions in expressing various semantic content categories and the content-form interface in children with language difficulties. Two Mandarin-speaking children with language difficulties were referred by a local therapist and their language samples were obtained and analysed. The transcribed data were compared with the results from children with typical development in study one, to find out their strengths and weaknesses. The study emphasizes the role of semantic content category, as well as content-form interface in enriching the language profiles and guiding intervention directions for children with language difficulties (see Chapter four).

After presenting the main studies and their results, chapter five provides a summary of the thesis's key findings. This includes the acquisition trajectory of semantic content category

in typically developing children, the content-form interaction in the acquisition of temporal markers, and case studies on manifestations of disruptions of semantic content categories and content-form interface in children with language difficulty. The importance of considering the content domain and the content-form interface in the studies of child language disorder is suggested. Clinical implications for the intervention of children with language disorders are also illustrated. Finally, the significance and limitations of the research are acknowledged, and directions for future research are suggested.

Chapter Two

Establishing the Database for Investigating Content-Form Interface in Mandarin-Speaking Children

2.1 Introduction

Children explore the world and learn to use language to communicate in their early years. The acquisition of early language has attracted much attention for many decades. Among the three major domains of language – content, form and use (Bloom & Lahey, 1978), the acquisition of form, as well as the content domain with vocabulary and semantic features, have been focused on predominantly by previous studies (e.g., Brown, 1973; Lieven et al., 1992; Scarborough et al., 1991). With growing emphasis on the importance of semantic content in the study of child language (e.g., Bloom, 1970; Hsu and Bishop, 2020; Mok & Kipka, 2009), the research gap in studying language acquisition in the domain of content deserves to be better investigated. This study aimed to investigate the acquisition of various semantic content categories (Lahey, 1988) in Mandarin-speaking children, in order to fill the research gap by contributing additional information from the semantic perspective. To achieve this goal, a child database utilizing naturalistic language data, with syntactic annotations with part of speech and semantic annotations with semantic content category was established to investigate the above issues. Both traditional language measures and the acquisition trajectories of semantic content category in young children were examined. In addition, the potential use of the database in examining the interaction between language content and form in the acquisition process was also explored. A brief introduction to various semantic content categories, followed by a review of previous studies on the acquisition of semantic content categories is given below.

2.2 Literature Review

2.2.1 Semantic Content Category in Young Children

Semantic content is one component of language that concerns with the meaning of information, including the concepts and themes that language expresses. Young children typically express contents that reflect their immediate environment and experiences (Bloom & Lahey, 1978). A wave of crosslinguistic studies in the 1970s has proposed that young children, regardless of their native language, exhibit common topics and ideas in their early expressions (e.g., Bowerman, 1973; Brown, 1973). Similarly, Slobin's (1985) crosslinguistic acquisition studies also echoed the restricted set of meanings expressed by young children. Other studies also reported cross-culturally common aspects of semantic relations (Raybeck & Herrmann, 1990) and antonymic meanings (Raybeck & Herrmann, 1996). As such, Lahey (1988) identified 21 semantic content categories in the production of young children according to the general kinds of meaning they share. In general, Lahey (1988) had classified these semantic content categories into four main classes. The first one represents information about objects or classes of objects (*existence*[2]*, non-existence*). The second one conveys the idea of action and events (*action, locative action*). Next, states of objects and events are also represented (*state, locative state*). Lastly, some semantic content categories express relations between objects, or between objects and events in a simple manner (*reject, denial, recurrence, possessive, attribute, quantity, notice, dative*), as well as the more complex relations within a single event or with another event (*additive, temporal, causal, adversative, epistemic, specification, communication*). These semantic content categories allow children to express most of the things and events around them, and are listed in Table 2.1 with examples.

---

[1] For ease of reference, all semantic content categories are italicized in the thesis.

Table 2.1. List of semantic content categories with examples, modified from Lahey (1988)[3]

No.	Semantic content category	Abbreviation	Example with Chinese translation
1.	*Existence*	exist	汽车 'car'
2.	*Recurrence*	recur	再 'again'
3.	*Nonexistence*	nonE	没有 'none'
4.	*Rejection*	rej	别 'do not'
5.	*Denial*	den	不是 'is not'
6.	*Attribution*	attri	大的 'big'
7.	*Possession*	poss	我的 'My'
8.	*Locative Action*	locA	去 'go'
9.	*Action*	act	睡觉 'sleep'
10.	*Locative State*	locS	上面 'above'
11.	*State*	state	饿 'hungry', 黑暗 'dark'
12.	*Quantity*	quan	许多 'many'
13.	*Notice*	noti	看到（他在跑）'see (him running)'
14.	*Dative[4]*	dat	给你（一支笔）'give you (a pen)'
15.	*Additive*	add	…和… 'and'
16.	*Temporal*	temp	刚刚 'just'
17.	*Causal*	caus	因为… 'because'
18.	*Adversative*	adver	但是… 'but'
19.	*Epistemic*	epis	怀疑（是他做的）'suspect (it is done by him)'
20.	*Specification[5]*	spec	穿红衣服的男孩 'the boy who wears in red'
21.	*Communication*	com	告诉（我一个秘密）'tell (me a secret)'

---

[3] It is believed that there may be some culturallyspecific content in Chinese. The culturally specific concern was considered but not included in the current study. This study intended to follow and extend the work of previous studies by adopting cross-cultural and cross-linguistic SCCs.

[4] In English, *dative* is mostly represented in SVO$_1$O$_2$ in which both direct and indirect objects are obligatory. In Chinese, the argument in *dative* expressions can sometimes be omitted (Chung & Gordon, 1998). In this thesis, the content expressed by dative is the main focus in which *dative* indicates the recipient or beneficiary of an object or action.

[5] Both content of *attribution* and *specification* involve describing features but their primary foci differ. *Specification* emphasizes differentiation and comparison between objects or events, highlighting their distinct qualities, while *attribution* primarily lists or describes associated features without necessarily emphasizing differentiation. The two SCCS were identified by considering the context and functions.

Adopting the framework of semantic content category, Brown (1973) analysed the first fifty words produced by children speaking English, Finnish and Spanish. Despite differences in the specific words used across languages, the semantic content categories were remarkably similar across different languages. Similar results were also obtained by Bowerman (1973). Besides, the studies of Stockman and Vaughn-Cooke (1982 & 1986) examined children's productions beyond single word level and provided evidence on the equivalent set of semantic content categories among young African America children, for which forms are mapped onto the same underlying concepts, in regardless of culture and whether standard forms are used.

On the other hand, cultural-specific meaning was also found in the semantic domains. Herrmann & Raybeck (1981) conducted a study on the expression of emotion among Ifugao children in the Philippines between the age of 2;0 and 3;6. Results showed a greater degree of emotional expressiveness and sensitivity towards negative emotions, such as anger and sadness, compared to positive emotions such as happiness. This may be related to the cultural emphasis on self-control and emotional restraint in Ifugao society, which encourages the suppression of overt displays of positive emotions. Russell (1991) also reported differences in the emotion categories across different cultures and languages, with the basic ones being more universal and the subordinate ones being more specific.

To the best of our knowledge, no previous studies have been done on the acquisition of semantic content categories by Mandarin-speaking young children. Despite a few previous studies have suggested cultural influences in particular meanings like emotion (e.g., Herrmann & Raybeck, 1981; Russell, 1991), the major studies appeared to indicate that the above-mentioned semantic content categories were found in early child language in general (e.g., Bowerman, 1973; Brown, 1973). Therefore, the current study adopted this specific set of semantic content categories and investigated their acquisition in Mandarin accordingly.

2.2.2 Acquisition of Semantic Content Categories

There are a few studies examining the acquisition of semantic content categories by English-speaking young children. The study conducted by Stockman and Vaughn-Cooke (1986) investigated the order of acquisition of various semantic content categories with twenty-two standard and nonstandard English speakers aged 18 to 54 months. Results indicated that *existence* and *action* were the earliest acquired semantic content categories and were the only ones productively produced by every child. After that, the semantic content categories of *locative action*, *state*, *locative state, negation*, *possession*, *attribution*, *notice and recurrence* also emerged, whereas semantic content categories of *additive*, *causal* and *epistemic* were acquired last. This study did not only present preliminary acquisition trend of various semantic content categories in young children, but also indicated that both standard and nonstandard speakers of English shared the same semantic base in early language although the forms used were different.

Bloom (1991) also carried out a series of longitudinal studies to investigate the acquisition of semantic content categories of children aged 24 to 38 months by using naturalistic language samples. The expressions of various semantic content categories related to early sentences and complex sentences were examined respectively. First, looking at the earliest appeared semantic content categories of simple sentences, contents *of existence*, *non-existence* and *recurrence* were found to emerge before *action*, *locative state* and *locative action*. Specifically, among the semantic content categories related to verb, it was found that *action* emerged first, followed by *state* and then *locative action*. Finally, *dative* and *specification* came along after the above basic verb relations. The emergence of *possessive* and *attribution*, on the other hand, varied among the participants. Apart from the above, a specific order of acquisition for semantic content categories of negation was also suggested, with *non-existence* emerged before *rejection*, while *denial* emerged last.

23

Next, eight semantic content categories associated with complex sentences were observed in the samples, including *additive*, *temporal*, *causal*, *adversative*, *specification*, *epistemic*, *notice* and *communication.* Among those involving the use of conjunctions, the acquisition followed the sequence: *additive > temporal > causal > adversative*. On the other side, *epistemic*, *notice* and *communication* were observed to emerge after the above, while the content category of *specification* was infrequent in the language samples of all participants.

2.2.3 Factors affecting the Acquisition of Semantic Content Categories

The above acquisition order of semantic content category allows us to understand how content affect language acquisition in young children. Proposed by Bloom (1991), this relative order of acquisition is affected by two major factors in general: syntactic complexity and cognitive complexity. From the syntactic view, semantic content categories relate to complex sentences in general appeared later than those associated with simple sentences (e.g., *existence* and *possessive* emerged before *causal* and *specification)*. When considering the semantic content categories expressed through verbs, it is observed that when expressing *locative action*, which typically involves describing places, longer and syntactically more complex utterances are utilized compared to utterances that solely describe motion or location alone in relation to *action* or *locative state*. Likewise, the expressions of *dative* content, which describe two different kinds of relations between recipients and affected objects, are also syntactically more complex than utterances expressing only *action* and *locative action*, and thus emerged later. On the other hand, regarding semantic content categories associated with complex sentences, those involving the use of syntactic structures with conjunctions (i.e., *additive*, *temporal*, *causal*, *adversative*) generally appeared earlier than those related to more complex complementation (i.e., *epistemic*, *notice*, *communication).* Semantic content category related to relativization *(*i.e., *specification)* is regarded as more complex syntactically and thus emerged last*.*

Cognitive complexity was suggested to be another factor accounting for the acquisition order of various semantic content categories according to Bloom (1991). Children learn that objects exist by acting in ways that make them disappear and recur in the sensorimotor period (Piaget, 2000). This relatively simple object permanence explains the earliest emergence of *existence*, *non-existence* and *recurrence.* Then, children learn that objects can be acted upon and located in space, as reflected in the emergence of verb relations including *action*, *state, locative state* and *locative action*, which came immediately after the previous object knowledge. In particular, *locative action* representing the dynamic concept of the movement of objects between two places, usually involving source, path and goal, is more complex cognitively and appears later than *action* and *locative state*, which are static in nature. Likewise, the expression of *attribution* requires discrimination among similar objects, and thus higher-level cognitive processes of categorization and seriation, whereas *action* only involves comparatively simpler sensorimotor schemas in its production. Therefore, *action* preceded *attribution* in the acquisition. Finally, the involvement of a cognitively more complex symbolic referent in *denial* may also partially account for its later acquisition than *non-existence* and *rejection*, which negate more concrete objects or actions respectively.

The principle of cognitive complexity also applies to the acquisition among semantic content categories associated with complex sentences. Bloom (1991) suggested that there is a progressive increase in the cognitive complexity among the acquisition of *additive, temporal, causal* and *adversative.* Firstly, *additive* simply represents the indefinite joining of two events and is regarded as relatively simple. *Temporal* conveys the relations of two or more events with designated temporal sequences, while *causal* relations express the antecedent and consequence of events on top of the additive and temporal concepts. *Adversative* appears to be the most complex one, which involves additive, temporal and sometimes, causal concepts, as well as expressing the new meaning of opposition which involves the cognitive process of comparing

and contrasting. As a result, the cumulative cognitive complexity among these semantic content categories tends to explain their corresponding status in the acquisition process.

To achieve the goal of investigating the acquisition of semantic content categories in Mandarin-speaking children, the acquisition trajectory reported in English will be used as a reference. In light of the above factors of syntactic and cognitive complexity, it is predicted that the acquisition of semantic content categories among Mandarin-speaking children basically follows the trajectory reported in English, with some minor differences due to the syntactic properties of Mandarin and cultural-specific factors.

2.3 The Current Study

In sum, while the majority of previous studies on child language acquisition had been focusing on the study of form or lexical semantics, relatively few studies examine the content domain beyond lexical level in early language acquisition. Notably, there is currently no study investigating the acquisition of semantic content categories among Mandarin-speaking children. As a result, the current study aimed to establish the Corpus of Mandarin Child Language (CMCL), a corpus with naturalistic language samples, for the study of the acquisition of semantic content categories in Mandarin-speaking children. Particularly, a specially designed protocol for eliciting various semantic content categories was adopted and the language samples were annotated with both part-of-speech and semantic content category. The following research questions were addressed in the study:

1. What is the acquisition trajectory of various semantic content categories among Mandarin-speaking children?

2. What are the similarities and differences in the acquisition trajectories of semantic content category between Mandarin-speaking and English-speaking children?

To address the concern on the applicability of the language data in CMCL, the measures on utterance length and lexical diversity of the children's productions were also obtained. It was expected that the age differences in these traditional measures from previous studies would be replicated in the current study. On the other side, to avoid the issue that children's language productions were dictated dominantly by the topics initiated by the interviewers, a standardized protocol was used to minimize the influence of different interviewers, as well as to create equal opportunities across children to elicit all the intended semantic content categories specifically. It was also predicted that the acquisition trajectory of semantic content category among Mandarin-speaking children would generally follow the sequences suggested in Bloom (1991). Syntactically, semantic content categories related to syntactically simple sentences would appear earlier than those related to syntactically complex sentences. Similarly, semantic content categories associated with lower cognitive complexity would also precede those cognitively more complex ones in the acquisition. However, one can anticipate slight variations arising from the unique linguistic and cultural considerations inherent to Mandarin. It is expected that subtle differences will exist in the acquisition of certain semantic content categories. For instance, as a result of the influence of Confucian teachings in traditional Chinese culture (Xiao, 1999), Chinese children are taught to exhibit greater obedience and may therefore demonstrate fewer content of reject. This current study not only sheds new light on the understanding of language acquisition in the semantic domain, but allows the investigation of the content-form interface and contributes to validating different theories in the early language of children. Moreover, clinical implications for the intervention of children with language disorders are also provided.

2.4 Methods

2.4.1 Participants

The Corpus of Mandarin Child Language (CMCL) was established with 82 native Mandarin-speaking children aged 25 to 60 months (48 boys and 34 girls) recruited from early education centres and kindergartens in Shenzhen and Guangzhou, China. According to their caregivers, none had any sensory or intellectual disabilities or language problems. All participants were divided into three age groups by one-year intervals, i.e., 25 to 36 months, 37 to 48 months, and 49 to 60 months. Information for these subjects is shown in Table 2.2.

Table 2.2 Subject information of CMCL

Age group	n	Age range (months)	Mean age (months)	n male	n female
1	19	25-36	30	13	6
2	41	37-48	42	24	17
3	22	49-60	55	11	11

2.4.2 Language Sample Collection and Transcription

Participants' language samples were elicited individually in a quiet room by speech therapists, speech therapy students and research assistants who had received prior training. First, a warm-up period with a doll set or train set was included to build rapport between the child and the researchers before conducting the three tasks of taking actual language samples (i.e., freeplay with toys, storytelling with pictures, and conversation). The semi-spontaneous speech of each child was then collected through one-to-one interactions with the examiner on three tasks following the standardized procedures in CMCL. Identical sets of toys, including a cooking set, food, utensils, puppets and a mystery bag, were provided during each 20-minute

freeplay session. Each child was encouraged to play and communicate with the examiner. Importantly, the play was specially designed to provide a scenario for eliciting different semantic content categories. Open-ended questions, parallel play and parallel talk were employed to facilitate the child's language production. A color Cookie Theft picture (Goodglass & Kaplan, 1972) and a set of four-card stories were then provided to elicit each child's narrative speech, which lasted for about five minutes. Finally, the examiner initiated a five-minute talk about daily life according to the child's interests and experience. A Peppa Pig storybook was also presented to provide topics for the chat (e.g., picnicking and favourite cartoon characters).

All sessions were audio- and video-recorded and all utterances produced by both the participants and interviewers were transcribed orthographically. Pauses of more than two seconds, intonation contours and speaker turns were used to determine the utterance boundaries (Klee & Fitzgerald, 1985). All the utterances produced by the participants were then analyzed by the trained speech therapy students and research assistants. However, following Crystal et al. (1989), self-repetition, unintelligible utterances, and incomplete utterances that did not reflect the children's language ability were not analyzed. The physical context was provided by the descriptions of the events and the actions of participants, while the linguistic contexts were shown with the examiner's utterances.

To ensure transcription accuracy, the transcribers, all holding a degree in linguistics, were extensively trained prior to the beginning of the study. Language samples obtained from a pilot study were transcribed by all transcribers together and any discrepancies were resolved through discussion. A manual with a detailed set of transcription guidelines was eventually agreed upon. The audio- and video-taped samples also allowed multiple viewing when necessary. Finally, several transcripts independently transcribed by each of the transcribers were compared using the RELY function in CLAN (Codes for Human Analysis of Transcripts)

to check for transcription consistency. The percentage of the overall match of words between the transcribers was 95.4%.

2.4.3 Database

All the above data were imported to the software Filemaker Pro which provides an interface to display the content of CMCL. The orthographic transcription of child utterances formed the first main layer in the database. Every single utterance produced by the child and interviewer was represented as one entry. Next, words within each utterance were identified according to the principles of boundedness, expandability, versatility proposed by Zhu (1982). For each word, tags on phonological form using *pinyin*, English translation and part of speech were added with reference to the MOR databank provided by Child Language Data Exchange System (CHILDES; MacWhinney, 2000). Words that could not be found in the 'zho' dictionary in the MOR databank were identified and added manually. The newly added words mainly included names of specific people, places, cartoon programme and cartoon characters. Specific food items (e.g., 甜甜圈 'donut') and kitchen utensils (e.g., 蒸锅 'steamer'), as well as reduplication of words (e.g., 鸡鸡 'chicken' (noun); 神秘秘 'mysterious (adjective)) were also added.

Finally, the utterances were annotated with 21 semantic content categories based on the descriptions in Lahey (1988), to form the third semantic tier. Assignment of semantic content categories was done by trained speech therapist students. Notably, a particular semantic content category can be encoded with different syntactic forms (e.g., content of *temporal* can be expressed with aspect marker, temporal adverb and temporal noun by Chinese-speaking young children (Tse et al., 2012; Zhou, 2004)), whereas a particular form can also be used to encode different semantic content categories (e.g., the phrasal structure 公园的树木 represents *location*, 红色的杯子 represents *attribution*, while 爸爸的头发 represents *possession*). In this

study, the semantic content categories associated with different utterances were annotated based on the participants' intended meaning determined according to the corresponding physical and language contexts shown in the video. After all the semantic coding had been done, 10 percent of the language samples were randomly selected and independently coded by a second rater to develop inter-rater reliability. A relatively high agreement of 93.0% of semantic content category coding was achieved between raters. A sample layout of the database is shown in Figure 2.1.

Figure 2.1. Layout of CMCL in Filemaker Pro

2.4.4 Data Analysis

Data were extracted from the CMCL for further analysis. Each child's major utterances except the deviant ones were used to calculate the mean length of utterance in word (MLU), following Cheung's (1998) and Zhu's (1982) procedures. Traditional measures on lexical

diversity including number of different words produced (NDW), total number of words produced (TNW), type-token ratio (TTR) and number of different open/closed class words were examined. In addition, vocD (Malvern & Richards, 2002) was also calculated automatically by running the MOR command in CLAN. Finally, the usage and acquisition of different semantic content categories were also analyzed.

2.5 Results

There are altogether 13, 630 utterances produced by 82 children from the age of 25 to 60 months (mean age = 43 months, 48 boys and 34 girls). Among all the child utterances, only 10,643 were analyzed in CMCL, while the remaining 2987 utterances (22% of all child utterances) including short responses, repetition, incomplete or unintelligible utterances were not analyzed. Results on utterance length and lexical diversity, as well as the usage and acquisition of different semantic content categories were illustrated as follows.

2.5.1 Utterance Length and Lexical Diversity

Table 2.3 presents the descriptive statistic information of children's MLU, NDW, TNW, TTR and vocD across different age groups. To investigate the effect of age on sentence length, a one-way analysis of variance (ANOVA) was conducted. Results showed a significant effect of age on MLU (F (2,79) = 11.02, p < .001). Post-hoc analyses using the Bonferroni test at a significance level of .05 indicated that the two year olds possessed lower MLU than the three year olds, but the difference between the three year olds and the four year olds was not significant.

Table 2.3 Different language measures of child utterances across age groups

	Age group 1			Age group 2			Age group 3		
	Mean	S.D.	Range	Mean	S.D.	Range	Mean	S.D.	Range
MLU	2.73	0.55	1.74-4.3	3.61	0.82	2.26-5.86	3.78	0.77	2.14-5.63
NDW	113.61	27.46	64-179	168.26	40.45	102-265	163.05	45.26	108-258
TNW	507.78	186.32	192-883	679.14	272.83	216-1446	537.09	249.7	277-1106
TTR	0.24	0.06	0.14-0.38	0.27	0.07	0.15-0.50	0.33	0.07	0.20-0.46
TNU	136.39	46.02	36-193	142.5	42.22	62-241	116.59	40.69	62-206
vocD	33.5	8.53	15.9-47.0	41.1	12	16.0-64.7	48.3	6.93	34.6-60.2

Note. MLU = Mean length of utterances in word; NDW = Number of different words;

TNW = Total number of words; TTR = Type token ratio; TNU = Total number of utterances

Regarding lexical diversity, a one-way ANOVA was conducted to analyze the effect of age on each of the measures of NDW, TNW, TTR and vocD. Results showed a significant age effect on NDW (F (2,79) = 15.58, p < .001), TTR (F (2,79) = 9.41, p < .001), and vocD (F (2,79) = 18.35, p < .001). Post-hoc Bonferroni test at a significance level of .05 revealed that the observed NDW was lower in the two year olds than the three year olds, while the three year olds also possessed lower TTR than the four year olds. On the other hand, lower vocD was also observed in both the two year olds and the three year olds when compared with their older counterparts respectively. Pearson's correlation was also conducted to investigate the relations between sentence length and various measures of lexical diversity. Table 2.4 presents the correlations among children's MLU, NDW, TNW, TTR and vocD. The results indicated that MLU is significantly correlated with NDW (r = .662, p < .01) and TTR (r = .662, p < .01), but not with vocD (p > .05).

Table 2.4 Correlation between children's mean length of utterances, number of different words, total number of words, type-token ratio and vocD

Variables	1	2	3	4	5
1. Age	-				
2. Mean length of utterance	.332**	-			
3. Number of different words	.313**	.662**	-		
4. Total number of words	-.066	.509**	.807**	-	
5. Type-token ratio	.513***	.662**	1.0**	.807**	-
6. vocD	.480***	.144	.145	-.081	.334**

Note. * indicates $p < .05$; ** indicates $p < .01$; *** indicate $p < .001$

The number of different open class words (nouns, verbs and adjectives) and closed class words produced by children in each age group was shown in Table 2.5. Results of correlations (Pearson's r), as shown in Table 2.6, indicated that children's age is significantly correlated with number of different nouns produced ($r = .415, p < .01$), number of different verbs produced ($r = .282, p < .01$), and number of different adjectives produced ($r = .451, p < .01$), but not with number of different closed class words produced ($p > .05$).

Table 2.5 Average number of different lexical items across age groups

	Age group 1		Age group 2		Age group 3	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Noun^	28.28	10.86	47.5	12.67	47.77	15.57
Verb^	21.78	7.12	34.05	9.86	32.36	9.51
Adjective^	8.06	3.32	17.19	6.33	17.95	5.71
Closed class words^	55.5	10.42	69.52	16.02	64.95	18.68

Note. ^Number of different items in that particular lexical class

Table 2.6 Correlation between children's age and number of different lexical items.

Variables	1	2	3	4	5
1. Age	-				
2. Noun^	.415**	-			
3. Verb^	.282*	.775**	-		
4. Adjective^	.451**	.748**	.729**	-	
5. Closed class words^	.103	.792**	.805**	.727**	-

Note. * indicates $p < .05$; ** indicates $p < .01$

^ Number of utterances expressing that particular lexical item

2.5.2 Semantic Content Categories

Table 2.7 presents the average number of different semantic content categories across age groups. Results of correlations (Pearson's r) indicated that the number of different semantic content categories is significantly correlated with age ($r = .269$, $p < .05$) and MLU ($r = .332$, $p < .001$). Notably, the semantic content category of *specification* was absent in all age groups.

Table 2.7 Number of different semantic content categories across age groups

	Age group 1			Age group 2			Age group 3		
	Mean	S.D.	Range	Mean	S.D.	Range	Mean	S.D.	Range
Number of different semantic content categories	14.5	2.2	9-19	16.5	1.90	13-20	16.4	1.40	14-19

Table 2.8 Token count of 21 semantic content categories across age groups

		Age group 1		Age group 2		Age group 3	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
1.	Existence	108.63	41.97	136.37	52.46	118.59	54.29
2.	Recurrence	3.11	4.92	2.41	2.55	1.50	2.15
3.	Nonexistence	6.16	2.83	6.90	5.65	3.36	2.46
4.	Rejection	3.11	2.62	3.32	4.03	1.14	1.75
5.	Denial	9.42	6.85	13.51	5.41	12.68	4.99
6.	Attribution	31.58	20.27	62.39	28.74	60.82	37.66
7.	Possession	4.53	4.31	6.05	4.98	8.91	7.23
8.	Locative Action	4.11	3.21	8.24	6.44	5.50	5.15
9.	Action	63.11	27.41	79.76	36.61	66.77	32.49
10.	Locative State	11.89	7.48	12.20	6.08	11.23	5.94
11.	State	15.32	8.87	21.88	12.38	21.59	12.20
12.	Quantity	7.58	5.81	12.24	9.49	11.27	8.48
13.	Notice	0.63	0.83	1.44	2.34	0.77	1.31
14.	Dative	2.05	1.84	3.00	4.06	2.77	2.52
15.	Additive	2.89	4.03	4.54	5.02	7.36	5.06
16.	Temporal	7.89	6.75	19.15	16.81	18.27	14.65
17.	Causal	1.26	1.73	2.85	3.29	3.86	5.44
18.	Adversative	0.11	0.32	1.54	2.31	1.36	1.84
19.	Epistemic	0.16	0.69	0.80	1.52	0.36	0.73
20.	Specification	0.00	0.00	0.00	0.00	0.00	0.00
21.	Communication	0.16	0.50	1.02	1.54	0.59	1.10

To investigate how children acquire different concepts within specific linguistic domains, a commonly employed approach is the utilization of a 90% criterion (e.g., Brown, 1973; So & Dodd, 1995). According to this criterion, a concept is considered acquired by the majority of children in a particular age group when 90% of them demonstrate correct usage of that concept during elicitation experiments. For example, researchers have examined the order of morpheme acquisition by looking at the age at which 90% of the children were reported to produce the morphemes (e.g., Brown, 1973; de Villiers & de Villiers, 1973). Likewise, a 90% criterion has also been used in research documenting the acquisition trajectory of different

phonemes among young children (e.g., Crowe & McLeod, 2020; So & Dodd, 1995; Zarifian et al., 2015). To investigate the age of acquisition for different semantic content categories, a 90% criterion was therefore adopted. A semantic content category was regarded as acquired by the particular age group if 90% of the participants in the group produced the semantic content category at least once in the sample collected. Table 2.8 summarizes the semantic content categories acquired by participants from 2 to 5 years old. Among the 2-year-old participants, nine semantic content categories were acquired (*existence, non-existence, reject, attribution, action, locative state, state, quantity, temporal*). Notably, the semantic content category of *reject* reached 90% occurrence in this age group but declined and did not meet the acquisition criterion in the two older age groups. Three more categories (*denial, possession, locative action*) were added to the inventory of the 3 year olds while another three categories (*dative, additive, causal*) were further acquired by the 4 year olds. Six semantic content categories, namely *recurrence, notice, adversative, epistemic, specification* and *communication*, were not fully acquired even by the 5-year-old group.

Table 2.9 Occurrence of different semantic content categories by age group

	2 year olds	3 year olds	4 year olds
Existence	solid	solid	solid
Nonexistence	solid	solid	solid
Attribution	solid	solid	solid
Action	solid	solid	solid
Locative State	solid	solid	solid
State	solid	solid	solid
Quantity	solid	solid	solid
Temporal	solid	solid	solid
Reject	solid	dashed	dashed
Denial	dashed	solid	solid
Locative Action	dashed	solid	solid
Possession	dashed	solid	solid
Additive	dashed	dashed	solid
Causal	dashed	dashed	solid
Dative	dashed	dashed	solid
Recurrence	dashed	dashed	dashed
Notice	dashed	dashed	dashed
Adversative	dotted	dashed	dashed
Epistemic	dotted	dashed	dashed
Communication	dotted	dashed	dashed
Specification	dotted	dotted	dotted

Note. A dotted line indicates that the semantic content category had <20% occurrence, a dashed line indicates 20-89% occurrence, and a solid line indicates ≥90% occurrence.

2.6 Discussion

The current study reported a database, the CMCL, that documents the language samples obtained from Mandarin-speaking children aged between 2 to 5 years old. The orthographically transcribed language samples were tagged with their corresponding part of speech and semantic content category. Before discussing the major focus of the acquisition pattern of semantic content categories, typical measures of utterance length and lexical diversity were obtained to ensure the representativeness of the content of the corpus. Results indicated that the effect of age on sentence length and lexical diversity from previous studies was replicated. Details were elaborated on in the following section.

2.6.1 Utterance Length and Lexical Diversity

In the current study, it was observed that shorter utterance length was associated with the younger age group in general. The increase in utterance length moving from the two year olds to the three year olds, as reflected in the MLU measure, is consistent with reports of previous studies on both English (Miller, 1981; Moyle et al., 2011; Paul, 2000; Rice et al., 2010) and Chinese children (Cheung, 1998; Jin & Jin, 2008; Klee et. al., 2004; Wu, 2020). On the other hand, observations of the 4 year olds appeared to indicate a plateau in sentence length in this age group and replicated previous findings that MLU may be less sensitive in capturing the progress in language abilities of children beyond the earliest stage of language acquisition (e.g., Klee & Fitzgerald, 1985; Smith & Jackins, 2014).

Regarding the lexical diversity among children, NDW (especially for the open class words) and TTR, appeared to reflect some age effects in the current study. However, it is noteworthy that the results also demonstrated positive correlations between NDW and TTR with MLU, but not between vocD and MLU. This echoed previous findings that suggested the

possible confounds of NDW and TTR in representing the vocabulary skills of children, as both of them may be affected by sentence length (Klee, 1992). To this end, the current findings tended to support the notion that vocD is more suitable for measuring lexical diversity in the context of Mandarin (Zhang & Zhou, 2020). Besides, vocD also appeared to be a more sensitive measure in reflecting the progress of lexical diversity across age groups, as significant differences were found between both the 2-year-old and 3-year-old groups, as well as between the 3-year-old and 4-year-old groups. In sum, similar to the measure of sentence length, the increase in lexical diversity across age groups, as indicated in the measure of vocD, is also consistent with previous results with English-speaking (Miller, 1991; Owen & Leonard, 2002) and Chinese-speaking children (Jin & Jin, 2008; Klee et. al., 2004; Wu at al., 2019; Zhang & Zhou, 2020). Generally, the results obtained from the CMCL replicated the typical language acquisition pattern from previous studies, which confirm that the procedures we applied in data collecting language samples provide a valid data source for studying language acquisition in young Mandarin-speaking children.

Open class words primarily convey the concrete content of the sentences whereas closed class words are usually more related to the grammatical aspects of sentences, and include relatively few members (Weber-Fox & Neville, 2001). It was observed that young children produced proportionally more open class words than closed class words in a previous study using parent questionnaires (Klintfors et al., 2009). In the current study, however, the productions of the open class words and closed class words were more comparable among the 2 year olds. The standardized elicitation probes used in the current study may be one possible reason that accounts for the different observations compared with previous literature. Given the utilization of standardized elicitation probes that primarily emphasized here-and-now contexts in the present study, it is expected that the 2 year olds, who possess greater proficiency in generating contextualized expressions, may exhibit a higher inclination to produce

exclusively the open class words targeted in the elicitation probes. As a result, they may generate only a limited number of such words. In contrast, closed class words encompass aspect markers, numerals, classifiers, determiners, pronouns, sentence final particles, and prepositions. Among these, certain closed class words such as aspect markers and sentence final particles hold significant prominence in Mandarin and are extensively utilized by young children (Erbaugh, 1992; Fang & Hengeveld, 2022). This may also lead to the production of apparently more closed class words among the 2 year old, resulting in a relatively even ratio between the number of open and closed class words observed.

The results from the current study indicated consistent findings on the traditional measures like MLU and vocD in the language acquisition of Mandarin-speaking children, thus confirming the representativeness of the corpus. In the following, the findings of the major focus on the acquisition of semantic content category in Mandarin-speaking children are discussed.

2.6.2 Acquisition of Semantic Content Categories and the Underlying Factors

Given the importance of the semantic domain in language acquisition mentioned before, the usage and acquisition of various semantic content categories were investigated to supplement the study of child language acquisition from the semantic perspective.

Results showed that the number of different semantic content categories increased with both age and sentence length. The increase in the unique semantic content categories thus offers a glimpse into the language acquisition of typically developing children from the semantic perspective along with the syntactic analysis, providing a more comprehensive picture of language acquisition. Moreover, by using the 90% criterion, the acquisition pattern of different semantic content categories in Mandarin-speaking young children were also investigated.

Among the 21 semantic content categories, nine of them, namely *existence, non-existence, reject, attribution, action, locative state, state, quantity* and *temporal*, were acquired early by the 2-year-old group. The early acquisition of these semantic content categories, except *temporal*, is generally coherent with previous studies (Bloom, 1991; Stockman & Vaughn-Cooke, 1986). The later acquisition of *temporal* content reported in Bloom (1991) may probably be due to their investigation of the temporal content related to the use of conjunctions only in that particular study. In contrast, aspect markers which occur in Mandarin-speaking children as young as 18 months (Zhou, 2004), were included in our study of the *temporal* content category. Besides, according to Hao et al. (2008), Mandarin-speaking children aged 17 to 30 months experience a substantial vocabulary growth, and are capable of expressing vocabulary of objects, people, places and actions, as well as using some quantifiers, descriptive words, and words about time. By relating these vocabularies to their corresponding semantic content category (i.e., content of *existence* can be expressed with objects and people; content of *locative state* can be expressed with place; content of *attribution* and *state* can be expressed with descriptive words; content of *action* can be expressed by action words; content of *quantity* can be expressed by quantifier; content of *temporal* can be expressed by time words and aspect markers), it is found that the acquisition trajectory on semantic content categories in our study generally replicates previous research findings.

Adopting Bloom's (1991) framework, the early acquisition of the abovementioned semantic content categories can be explained by the relatively simple cognitive complexity of the concept being coded. Johnston (1985) suggested that conceptual development is the prerequisite for semantic growth. Likewise, Slobin (1973) also proposed that the complexity of the concepts affects the acquisition of linguistic terms, in which more abstract and complex ideas are acquired later. As a result, the cognitive complexity of the underlying concepts inevitably affects children's semantic representations, and thus the acquisition of various

semantic content categories. Specifically, acquisition of the semantic content categories of *existence, non-existence* and *action* requires the sensorimotor knowledge related to object permanence and should therefore be acquired in the sensorimotor period before the age of two (Piaget, 2000). In addition, concepts of objects and events are perceptually easy as concrete referents are usually involved. Lahey (1988) also claimed that young children have possessed the basic knowledge of objects and events at an early age. Among the earliest acquired semantic content categories, *existence* and *non-existence* mainly represent objects, people and events; *locative state* is related to different locations; *action* conveys the meaning of motion which does not involve changes of location. Besides, *reject, attribution, state, quantity* and *temporal* mostly serve to provide additional information of the object, action and event. The mapping between this concrete sensorimotor and cognitively simple knowledge of single object and event with words can therefore be established at an early age, and is reflected in their naturalistic speech expression.

Alternatively, the acquisition of these semantic content categories may also be related to linguistic factors. As children's early vocabularies contain proportionally more open class words than closed class ones (Klintfors et al., 2009), it is not surprising to observe that some of the above semantic content categories which are mostly coded using open class words (e.g., *existence, action, attribution)* were acquired early. Furthermore, most of these semantic content categories can be coded with simple syntactic units, such as single words (e.g., *existence, non-existence, attribution, action, locative state, state*), simple noun phrases (e.g., '一个' [numeral + classifier] to code *quantity*) and simple verb phrases (e.g., '吃了' [verb + aspect marker] to code *temporal*). In line with Bloom's (1991) proposal on the effect of syntactic complexity, semantic content categories related to simple sentences were therefore acquired first.

Moving on to the 3-year-old group, three more semantic content categories, namely *denial, locative action* and *possession* were subsequently acquired. The emergence of these semantic content categories also followed the acquisition sequence reported in Bloom (1991) and Stockman and Vaughn-Cooke (1986) in general. The above acquisition is again considered to be related to the cognitive and syntactic complexity. Firstly, children progressively acquired the simple relations between objects and events on top of the knowledge in single objects and events (Lahey, 1988). To this end, the content of *denial* which negates attribute, identity and state of events (Chang, 1992), and the content of *locative action*, which expresses the change in locations including the source, goal and/or path, are used to represent the relations between objects and events. Similarly, the content of *possession* also expresses the relations between the owner or possessor and the entity. Acquisition of these semantic content categories representing the simple relational knowledge between objects and events tended to involve higher demand in cognitive processing, and therefore these semantic content categories emerged later than those acquired by the 2-year-old group.

The syntactic complexity associated with these semantic content categories may also play a role in the acquisition. To express the content of *possession* in Mandarin, it is proposed that the genitive marker '*de*' is the most common linguistic device being used (Shi & Zhou, 2018). Li (2004) also reported that the comprehensive use of possessive expressions with '*de*' was acquired by the age of three, which appears to be in line with our observations on the acquisition of *possession*. Meanwhile, this grammatical marking '*de*' is often regarded as a bigger challenge than the content words in the acquisition by young children (Huang et al., 2022). On the other side, the content of *locative action* may also utilize co-verbs or prepositional phrases to encode both the manner and path of the motion in Mandarin (Slobin, 2004). The unique thematic roles associated with different prepositions in Chinese may further impose difficulties in the acquisition of the corresponding constructions (Lau et al. 2023).

These specific language properties of Mandarin thus constitute a higher syntactic complexity in encoding both *possession* and *locative action,* and account for their acquisition beyond 2 year old in Mandarin-speaking children.

Next, the 4-year-old group further acquired the semantic content categories of *additive, causal* and *dative*. This again replicates the acquisition order reported in both Bloom (1991) and Stockman and Vaughn-Cooke (1986). The acquisition trend in this age group can also be explained according to the cognitive and syntactic complexity associated with the semantic content categories. First, both *additive* and *causal* represent specific relations of two events or states and should be cognitively more complex than those representing only one event. Besides, the content of *causal* additionally indicates a dependency between the events and gives the reason or result of the events, thus requiring more cognitive resources to process (Bloom, 1991). The above contributes an initial explanation for the acquisitions of *additive* and *causal* in this group of children.

On the other hand, acquisition of *dative* seems to be more related to the syntactic complexity. The coding of *dative* in Mandarin usually involves the use of different syntactic structures and prepositions with different themes (e.g., '送花给妹妹' [verb + noun + preposition + noun]). (Details about different syntactic structures and use of prepositions are elaborated in later section.) As mentioned before, Chinese prepositions are usually associated with specific thematic roles, which impose additional challenges on the syntax constructions (Lau et al., 2023). As a result, a more complex syntactic structure entailing the relations between recipients and the affected objects is usually involved with *dative* expressions. Similarly, the coding of *additive* and *causal* contents also requires the conjoining of phrases or clauses with conjunctions (e.g., '面包和鸡蛋' [noun + conjunction + noun] to code *additive*; '生病就看医生' [verb + conjunction + verb] to code *causal*). This seemingly higher syntactic

45

complexity provides further explanations to their acquisition in this later stage of early childhood.

Finally, the semantic content categories of *recurrence, notice, adversative, epistemic, specification* and *communication* were found not to be acquired by the 4-year-old group, using the 90% criterion. Once again, the late acquisition of the above semantic content categories, except *recurrence,* aligns with the order suggested by Bloom (1991) and Stockman and Vaughn-Cooke (1986). From the cognitive perspective, as noted in the previous section, an increment in the cumulative cognitive complexity among *additive, temporal, causal* and *adversative* is suggested. The content of *adversative* not only represents definite events with the highest cumulative cognitive complexity, but additionally denotes complex contrastive relationships beyond the here-and-now context (Bloom, 1991). The highest cognitive complexity of *adversative* thus explained its later acquisition. On the other hand, it is also proposed that the semantic content categories about how people think (*epistemic*) or talk (*communication*) about events, which involve implicit and non-transparent reasoning processes, may be complex for young children to interpret because it is relatively difficult for them to take the perspective of others (Evers-Vermeul & Sanders, 2011). In addition, some specific lexicons are also required in representing the definite relations of events in certain semantic content categories (e.g., perceptual verb to code content of *notice*; mental state verb to code content of *epistemic*; verb for conversation to describe what is to be express in the content of *communication*). The challenges in acquiring these specific and abstract words therefore pose extra difficulties in acquiring the corresponding semantic content categories.

Syntactically, in Mandarin, expressions of *adversative* usually require the use of conjunctions between clauses, and generally appear after those associated with simple sentences, but earlier than those related with more complex complementation (i.e., *epistemic*,

notice, communication). Semantic content categories related with complementation, subsequently, also appeared earlier than those related to relativization *(*i.e., *specification).* The higher syntactic complexity of these semantic content categories may therefore explain why they are not acquired by the 4-year-old children.

It is noted that the semantic content category of *specification* was absent at all ages in the current study, which is coherent with the infrequent occurrence in young children reported in Bloom (1991). According to Lahey (1988), *specification* is usually expressed to indicate a particular person, object or event in terms of their functions, places or activities. More specific relational knowledge is thus required for its production. In addition, expression of *specification* usually involves complex sentences with relativization (e.g., 穿红裙子的女孩在哭) (Bloom, 1991). According to Arndt and Schuele (2013), typically developing children usually start to form sentences containing relative clauses between the ages of 4 to 5 years of age and they continue to master the productions through the school-age year. It is therefore believed that this semantic content category only emerged at an older age, probably after the age of five due to higher demand in cognitive ability and syntactic knowledge.

Unexpectedly, the semantic content category of *recurrence* was not acquired by the age of five in the current study. Likewise, it is also interesting to observe that the occurrence of the early-acquired semantic content category of *reject* declined across ages and its occurrence did not reach 90% after the age of three. Both contents represent knowledge of a single object and event and possess relatively low cognitive complexity. Syntactically, both can be encoded with simple phrases (e.g., 再吃 [adverb + verb] to code *recurrence*; 不玩 [negative marker + verb] to code *reject*). With relatively simple cognitive complexity and syntactic structure, these contents are expected to be acquired early. One possible reason for the current findings is that the provided communicative context may not be obligatory enough to elicit the content

categories of *recurrence* and *reject*. In addition, the high occurrence of *reject* in the two-year-old group may also be related to the stage of 'trouble two' in children's development. It has been proposed that frequent noncompliance is common in toddlerhood and peaks during the second year (Alink et al., 2006; Tremblay et al., 2004). Most children then learn to regulate their behavior as they grow into preschool years and their noncompliance declines (Tremblay et al., 2003). Our findings appear to be consistent with this pattern of child development. On the other hand, it is also possible that some cultural-specific factors may affect the production of these semantic content categories. Accordingly, social harmony and fitting in with others are more encouraged in the Chinese context than in the Western ones (Wu, 1996). According to Xu and Farver (2009), Chinese children tend to exhibit regulated shyness, as a self-controlled form of social restraint to maintain harmonious group functioning and exemplifies self-regulation. Since *reject* and *recurrence* are mostly expressed when children intend to refuse or make additional requests on objects or events, these may not adhere to the social harmony in the Chinese culture. Similarly, it is also suggested that Chinese children are instructed to be more obedient due to the influence of Confucian teachings in traditional Chinese culture (Xiao, 1999). Older children may therefore tend to produce fewer negative expressions to reject, especially when interacting with our examiners who are all adults during the data collection. Considering all of the above, young Mandarin-speaking children therefore appeared to express fewer *recurrence* and *reject* contents in the current study.

To conclude, the acquisition trajectory of semantic content category in Mandarin-speaking-children aligns with the sequence suggested by Bloom (1991) for English-speaking children. Acquisitions in both languages are predominantly predicted by the associated cognitive and syntactic complexity, while specific linguistic properties in Mandarin and Chinese culture also seem to have a role in modulating the acquisition of particular semantic content categories.

2.6.3 Potential Usage of the Database in Investigating the Content-Form Interface in Language Acquisition of Young Children

According to Bloom (1991), young children acquire content and form together to lay the foundations of early expressions, and these two domains interact in a bidirectional manner to regulate the acquisition of early child language (Mok & Kipka, 2009). As mentioned, a certain form can be used to express various semantic content categories, while an individual semantic content category can also be expressed with different forms. On some occasions, children who fail to encode the concept with specific grammatical forms, may still be able to possess the semantic content category with the expressions of the non-standard or ungrammatical forms (e.g., 爸爸鞋鞋 'daddy shoe' with the absence of the genitive marker '*de*' to express *possession*). In view of the above, the importance of considering both language form and semantic content, as well as their interaction in child language study should not be underestimated.

The current study adopted language sample analysis, which allows the elicitation of various semantic content categories even among the youngest children by using a standardized protocol. The established CMCL not only provides important information on the acquisition of various semantic content categories in Mandarin-speaking children, but allows further investigation on the interaction between language content and form. By analyzing language samples annotated both syntactically with parts of speech and semantically with semantic content categories, the relations between content and form produced by Mandarin-speaking preschoolers were further investigated. In the following, the syntactic category of verb and the semantic content category of *dative* were used as examples to illustrate the usage of the database to explore the interaction between content and form in the language acquisition of Mandarin-speaking children.

2.6.3.1 Expression of Verb (Form) with Different Semantic Content Categories

In the past, a vast amount of research studies have been conducted to investigate the acquisition of early vocabulary by young children but conflicting results in the acquisition of nouns and verbs in early child language have been documented in the literature. Gentner (1982) proposed that nouns are universally acquired before verbs, while others found that verbs can also appear in children's earliest vocabularies (Ma et al., 2009). Indeed, Tardif (1996) argued that young Mandarin-speaking children produced more verbs than nouns in their naturalistic speech. In view of the above, the different semantic content categories encoded using verbs in the CMCL were explored in the current study. By searching utterances annotated with the part of speech 'verb' in the CMCL, it is found that verbs were used to express semantic content categories of *action*, *state*, *locative action*, *dative*, *notice*, *communication* and *epistemic*. Consider the proposed order of acquisition, *action* and *state* were the earliest semantic content categories to be acquired among all. *Locative action* and *dative* followed these two and were acquired by the 3-year-old and 4-year old children respectively. Finally, the content of *notice, communication* and *epistemic* were not acquired till the age of five. As a result, it will not be adequate to look at the surface form of the vocabulary alone, but to examine different semantic content categories being expressed by the form, in order to have a more comprehensive picture of early language acquisition. By classifying verbs according to the different semantic content categories being represented, it is possible to explain why some verbs appear earlier in age whereas others emerge later. It appeared that the acquisition of different lexical forms in young children is affected by the semantic content being encoded. With the syntactically and semantically annotated database, it is possible to investigate the content-form interface more thoroughly.

2.6.3.2 Expression of Dative (Semantic Content Category) with Different Forms

Next, the expressions of the semantic content category of *dative* with different forms were investigated. Utterances expressing the content of *dative* designate the recipient of an object or action with or without a preposition (Lahey, 1988). These utterances were extracted from the CMCL for further investigation of different lexical items and sentence structures produced by participants across ages. The lists of lexical items produced are shown in Table 2.9.

Table 2.10 Lexical items expressing dative content produced by each age group

Age group	Lexical items expressing dative content
1	给 1[i], 给 2[ii], 给 3[iii], 帮
2	给 1[i], 给 2[ii], 给 3[iii], 帮 ，喂，陪，让，请
3	给 1[i], 给 2[ii], 给 3[iii], 帮 ，喂，陪，让，请，跟，送，对

i. 给 1 serves as a verb (e.g., 给我一本书)

ii. 给 2 serves as a preverbal preposition (e.g.,给他送一本书)

iii. 给 3 serves as a post verbal preposition (e.g., 送给他一本书 or 送一本书给他)

(Li & Thompson, 1981; Liu, 2006)

The repertoire of lexical items expressing *dative* content was observed to be expanding across age groups. Older children were more capable of using a larger variety of lexical items, including verbs and prepositions, to express the content of *dative*. Considering the syntactic structures, it was found that the 2 year olds were able to produce sentences with double object construction (e.g., 我给他这个), prepositional construction (e.g., 给她穿上衣服) and serial verb construction (e.g., 奶奶帮它刷) in expressing the content of *dative*. On the other hand, the older children were capable of using pivotal construction (e.g., 我们要请你吃饭) in addition to the above sentence types. In addition to earlier findings that the content of *dative*

was only fully acquired by children in the oldest age group, these results indicate that older children were able to express the content with a larger variety of lexical items and syntactic structures.

From the above examples, by searching utterances annotated with verbs, it is possible to examine different semantic content categories being expressed. On the other hand, by searching utterances annotated with *dative* content, different lexical items and syntactic structures produced by participants across age groups can also be analyzed. Thus, the CMCL provides a useful and convenient tool for us to study the language acquisition of young children, from both syntactic and semantic perspectives. A preliminary platform for investigating this interaction between semantic content and syntactic form in the language acquisition process of Mandarin-speaking children is provided. It is further proposed that children acquire language in a two-dimensional way with various content-form interactions. When acquiring a particular form, children progressively broaden the semantic content categories being expressed by that form. Similarly, when acquiring a particular semantic content category, children also gradually expand the use of different lexical items and syntactic structures to represent that semantic content category.

In this study, the CMCL that documented the production of both content (semantic content category) and form (part of speech) was established. While currently available cross-sectional Mandarin corpora only possess syntactic annotations (e.g., Li & Zhou, 2008; Li & Zhou, 2015), the CMCL with specially designed elicitation procedures for semantic content category and the corresponding annotations, undoubtedly contributes potential values in language acquisition studies. By analyzing data extracted from the database, our findings not only provide insights into the acquisition trajectory of semantic content category in Mandarin-

speaking children, but also shed new light on the understanding of language acquisition in the content domains. It is further proposed that apart from the influence of cognitive and syntactic complexity, specific language properties in Mandarin and Chinese culture also seem to play a role in the acquisition. More in-depth investigation on the acquisition of particular semantic content categories in Mandarin is therefore inspired and supported by our preliminary findings. The CMCL will also be interesting for the broader research community and allow further research including the investigation of the content-form interface in the early language of children. It may further contribute to validating different theories in child language acquisition. Clinically, our findings inspire a potentially more comprehensive approach to profile the language ability of children with typical development, as well as those with language disorders. Assessment and intervention for children with language disorders can be planned with a balanced consideration between content and form (A more detailed investigation on the clinical implications of the content-form interface can be found in chapter four).

2.7 Concluding Remarks

The current study reported the findings on language acquisition in Mandarin-speaking children using a syntactically and semantically annotated database – CMCL. The replication of findings on sentence length and lexical diversity in typical language acquisition confirms the quality of the language sample data in the corpus for studying child language acquisition. Coherent findings in the acquisition of semantic content categories also appear to suggest that the acquisition trajectory of semantic content categories in Mandarin-speaking children mostly resembles that of English-speaking peers. Notably, the acquisition trend is predominantly explained by the cognitive and syntactic complexity, with additional influences from the language-specific properties and cultural-specific factors in Mandarin. In addition, with the tags in both part of speech and semantic content category, the CMCL not only contributes

additional perspective in studying child language from the semantic domain, but provides a useful platform for examining the interaction between semantic content and language form in early child language acquisition. Preliminary results suggested that the verbs expressing contents of *action* and *state* emerged earlier than verbs expressing other contents, such as *notice* and *communication*. Likewise, the content of *dative* expressed by older children also involved a larger variety of lexical items and syntactic structures than the younger ones. These findings subsequently imply significantly on the theoretical and clinical ground.

Chapter Three

Content-Form Interface in the Acquisition of Temporal Markers by Mandarin-speaking children

3.1 Introduction

Concepts of time are important constructs of human cognition. The ideas of past, present, and future serve as crucial mental frameworks for organizing our experiences and they are essential to individuals' expressions of the events taking place around them. Besides, expressions of time also play a crucial role in cultural transmission (Levine, 2024). In talking about time, speakers of different languages use forms in a range of specialized expressions. In English, for instance, inflectional morphology is employed extensively to express the internal time structure of events (aspect), as well as to indicate the location of events on a timeline of when speech is produced (tense). Furthermore, previous studies have found that the use of linguistic devices by children to express time varied according to the children's progress in the acquisition of more advanced time concepts (Weist, 1989). Mandarin, on the other hand, does not have morphological markers to express tense (Matthews & Yip, 2011) and instead relies on other linguistic devices to express time. While existing studies have indicated that Mandarin-speaking preschool children tend to use aspect markers (e.g., Li & Bowerman, 1998), temporal adverbs (e.g., Liang et al., 2019), and temporal nouns (e.g., Grant & Suddendorf, 2011) to express time, the acquisition of various time concepts by young children has not been fully elaborated. It is crucially important for early childhood educators and others, including speech therapists, to understand how children acquire the ability to refer to time using language. Therefore, this study investigated the acquisition of different temporal markers in expressing the *temporal* content by Mandarin-speaking children. The relationship between the acquisition of different time concepts and these markers was also examined, with the aim of opening a

window through which we can infer such children's development of the conceptual system of time (McCormack & Hoerl, 2017). Moreover, to illustrate the important role of the content-form interface in early language acquisition (e.g., Mok & Kipka, 2009), as proposed in study one, this particular semantic content category of *temporal*, was investigated thoroughly in the current study, to show how the acquisition of content and form affected each other.

3.2 Literature Review

3.2.1 Acquisition of Temporal Markers by Mandarin-Speaking Children

In Mandarin, the time structures of events can be conveyed via discourse, contextual resources, and linguistic devices (Huang, 2003; Li & Shirai, 2000; Smith & Erbaugh, 2005). In discourse, time is inferred from background knowledge and contextual information about an event that is shared between the speaker and the listener. Because young children have not yet mastered the discourse-pragmatic resources needed for dealing with time (Huang, 2003), they mostly rely on lexical devices to express it. Aspect markers (AMs), temporal adverbs (TAs), and temporal nouns (TNs) are the linguistic devices commonly used to express temporality by young Chinese-speaking children (Tse et al., 2012; Zhou, 2004).

Looking at AMs first, *aspect* refers to the internal temporal referents of a situation, which can be its beginning, its continuation, or its completion stages (Tang, 2016). Mandarin's four AMs are the perfective 了 *le*, experiential 过 *guo4*, progressive 在 *zai4* and durative 着 *zhe* (Li & Bowerman, 1998; Liu, 2015; Tang, 2016). The acquisition of Mandarin AMs has been found to occur in children as young as 18 months (Zhou, 2004). Generally, the perfective *le* is the first AM to emerge, followed by the progressive *zai4* and the durative AM *zhe*, while the experiential *guo4* is acquired last (Li & Bowerman, 1998).

TAs, such as 已经 *yi3jing1* 'already' and 马上 *ma3shang4* 'immediately', are also lexical devices commonly used to express time in Chinese. Semantically, TAs can be classified into three subtypes: past, present and future (Bi & Peng, 2002). Their use is widely agreed to emerge between the ages of 24 and 30 months (Liang et al., 2019; Zhou, 2004), but findings on the emergence of the different subtypes have been inconsistent. While most studies have reported that TAs related to the present emerge first, some have found that those related to the past appear second and those related to the future occur third (Bi & Peng, 2002; Zhu et al., 1982), although other studies have found that this order is reversed (Kong & Fu, 2004; Zhou, 2004).

Finally, in addition to the two more widely studied categories of temporal markers above, TNs such as 今天 *jin1tian1* 'today' and 分钟 *fen1zhong1* 'minute' also play important roles in denoting time points and event durations. Previous studies on the acquisition of duration TNs have shown that children aged up to six, despite possessing knowledge of the lexical categories and rank ordering of time words (Tillman & Barner, 2015), have an incomplete understanding of both their meanings and the absolute durations they represent (Shatz et al., 2010; Tillman & Barner, 2015). On the other hand, Grant and Suddendorf (2011) used parental-questionnaire data to study the production of 18 temporal terms, mostly TNs expressing a specific time point but also a few TAs and adverbial clauses, and found that non-specific temporal terms and those representing the present emerged the earliest. Surprisingly, however, TN acquisition in Mandarin-speaking children has received little scholarly attention. One possible reason is that TAs and TNs in Chinese can be hard to distinguish due to their morphological and lexical similarities (Biq & Huang, 2016). In addition, Chinese TNs can sometimes function as adverbials (Biq & Huang, 2016; Shi, 2016), which is likely why Zhu et al. (1982) lumped TAs and TNs together under the rubric of "temporal words." Similarly, Erbaugh (1992) classified TNs as "time adverbs" in a study on temporality acquisition. The

present research, in contrast, examined TNs independently of TAs due to their unique syntactic properties and the involvement of reference-time concepts in the Mandarin temporal system, as will be elaborated in a later section.

It is worth noting that most of the prior studies on expressions of time by children have primarily focused on a single category of temporal markers, for example, just AMs (Li & Bowerman, 1998), just TAs (Bi & Peng, 2002; Liang et al., 2019), and just TNs (Shatz et al., 2010; Tillman & Barner, 2015). One of the few studies to have examined the overall acquisition of multiple categories of temporal markers, by Tse et al. (2012), investigated the repertoires of AMs, TAs, and TNs among Cantonese-speaking children aged three to five. That study reported no significant change in AMs, but marked expansion in the repertoires of TAs and TNs as time went by. Zhou (2004), who studied the acquisition of all three categories of temporal markers by Mandarin-speaking children, reported that AMs were acquired earlier than TAs and TNs. Erbaugh (1992) also briefly described the dominance of AMs in Mandarin-speaking children's early years, followed by subsequent expansion of their inventory of TAs (and some TNs). However, the above-cited studies did not focus on the reasons for the acquisition trends that they had observed. A thorough investigation of the factors underlying the acquisition of all three categories of Mandarin temporal markers is therefore overdue.

3.2.2 The Influence of the Concept Development on the Acquisition between Different Categories of Temporal Markers

Weist (1989) suggested that speech time (ST), event time (ET), and reference time (RT) are concepts essential to people's development of a temporal system. Specifically, ST is the time point of a speech act, whereas ET refers to the time that an event occurred relative to ST. For example, in 'I watched a movie', the past tense is used to mark the ET, which is anterior to the ST. RT, on the other hand, refers to the temporal context established for the event in focus

and indicates the speaker's temporal vantage point (Zhang & Hudson, 2018). For example, in 'I watched a movie before three o'clock', the ET is relative to the RT 'three o'clock', which is different from the ST. According to Weist (1989), as their cognitive ability advances, children become more capable of developing time concepts that deviate from the here-and-now in four distinct stages, each marked by the use of different linguistic devices. In the first stage, children primarily focus on the here-and-now and locate the event in the deictic center of the speech act such that ST, ET, and RT all coincide. Before proceeding to the second stage, children begin to code the temporal contours of events, but the concept of ET is yet to be established. Then, in the second stage, the ET concept emerges and is progressively dissociated from ST, while RT remains unavailable. Children at this stage are capable of producing the aspect and tense contrasts of the same verb. Next, in the third stage, the RT concept emerges, and children demonstrate a temporal shift (i.e., they shift their perspective to a time other than ST). Nevertheless, RT, at this stage, unavoidably coincides with the temporal location of either ST or ET (McCormack & Hoerl, 2008). This restricted RT is anchored using TAs and temporal adverbial clauses. Finally, during the fourth stage, children demonstrate both a temporal shift and a seriation of time concepts in a "free" RT system; that is, they begin to use the past perfect tense and prepositions such as "before" and "after" to express ST, ET, and RT independently, and RT does not coincide with either ST or ET (McCormack & Hoerl, 2008). As children's time concepts affect their semantic representations of time, advancement in their knowledge of these time concepts inevitably affects their acquisition of time expressions.

To investigate the effects of the development of time concepts on the acquisition of temporal expressions, Weist et al. (1991) conducted a cross-linguistic study with 60 children aged two-and-a-half to six-and-a-half who spoke English, Polish, or Finnish. Results indicated that the children had more difficulty with RT temporal configurations during both comprehension and production tasks, regardless of which language they spoke. Nevertheless,

the Finnish children demonstrated a slightly different pattern from the other two linguistic groups when solving aspect and tense problems, probably due to the former's slower development in the comparatively more complex language system for time. On that basis, Weist et al. (1991) proposed that conceptual development places a universal constraint on the acquisition of time expressions, and the specific properties of individual languages also contribute considerable variation to that acquisition process.

In line with similar studies involving other languages, previous work on the acquisition of Mandarin temporal markers has mainly focused on a single category (e.g., Li & Bowerman, 1998; Liang et al., 2019), despite the possibility that combinations of such markers are involved in the free RT system. A recent paper by Li et al. (2022) represents a rare exception to this situation. The current study aims to further fill this gap by investigating how time concepts and all three categories of temporal markers are interrelated as a holistic system.

3.2.3 Semantic Factors Affecting Acquisition within Each Category of Temporal Markers

Liang et al. (2019) observed that children as young as two-and-a-half expressed certain examples of both past and future TAs, and that the variety of TAs that they commanded expanded across time. It has also been reported that a few "general" temporal terms were acquired before some "specific" ones (Grant & Suddendorf, 2011; Zhou, 2004). Similarly, TNs such as 'today', 'yesterday', and 'tomorrow' were used at earlier ages, and more accurately, than more distant temporal terms, including those for days of the week and months of the year (Grant & Suddendorf, 2011).

According to Clark's (1973) Semantic Feature Hypothesis, words are represented by multiple semantic features, and the more specific the meaning a word has, the more semantic features it will contain. Clark's (1973) hypothesis further holds that children do not acquire the

full meaning of any word at the beginning, but rather they learn words by continuously adding features to their lexical entries over time to differentiate between closely related words, and only eventually do they begin to acquire identifiable words. Under this hypothesis, the order of word acquisition is affected by the number and type of semantic features they contain. Previous studies have investigated the effects of semantic specificity on the acquisition of verbs and reported that verbs with more general meanings (and thus fewer semantic elements) were acquired earlier than those with more specific meanings (and more semantic elements). For example, Pinker (1989, p. 171) proposed that verb meanings arise from a set of specified semantic elements and that general verbs (e.g., be, have, go, do, make, put, give, take, and get) are acquired earlier than specific verbs. Children therefore need to identify the elements that differentiate specific verbs from general verbs. Similarly, Bloom (1991) claimed that children's early verb categories were semantically based, and that the specificity of verbs predicted their order of acquisition. Accordingly, the present study examined the effects of specificity on the acquisition of temporal markers and assumed that generic temporal markers had a privileged status in acquisition relative to specific markers.

Temporal remoteness has also previously been investigated as a factor potentially influencing the acquisition of temporal terms (Wagner, 2018). Young children's increasing ability to express events with increasing temporal remoteness as they aged was reported by Weist (1986); that is, they not only began referring to temporally proximal events at earlier points in their lives, but also referred to them more frequently than remote events thereafter. Similarly, Erbaugh (1992) claimed that Chinese children gradually extended their ability to describe events in the more remote past and future. Effects of temporal remoteness on the acquisition of temporal terms has also been evidenced among children who speak other languages, such as German (Szagun, 1979), English (Weist & Buczowska, 1987) and

Cantonese (Tse et al., 2012). As such, this study hypothesized that within each category of temporal markers, proximate markers would emerge at younger ages than distant markers.

3.3 The Current Study

Given that existing literature primarily focused on the acquisition of one single category of temporal markers, and most of the studies did not examine the possible factors affecting the acquisition trajectory they observed, the current study aims to fill the gap by investigating the acquisition between and within different temporal-marker categories. First, to investigate how Mandarin-speaking children acquire different temporal markers (form) to express the *temporal* content, Weist's (1989) model of temporal system was adapted to examine whether the availability of time concepts would affect the acquisition of Mandarin's three categories of temporal markers. The hypothesis was that various temporal markers would be employed for the children's transition out of the ST system (i.e., beyond stage one) and for anchoring ET and RT thereafter. It was expected that, as in Weist's (1989) first stage in which only the ST system exists, no temporal markers would be needed because the children would mainly focus on here-and-now events (e.g., 我吃饭 'I eat.'). Then, in the transition from the first stage to the second stage, AMs denoting the internal timeframes of events would emerge, prior to the emergence of the ET concept (e.g., perfective 了 in 我吃了饭 'I had a meal.'[6]). Next, the use of TAs would emerge, signaling the development of the ET concept in the second stage (i.e., the ET system), and they would be used to indicate the past/non-past timeframes of events (e.g., 已经 'already' in 我已经吃饭了 'I already ate.'). Additionally, TNs that represent time entities and denote the referents of time (Shi, 2016) would function to anchor separate RTs for events, and their

---

[2] For ease of reference, temporal markers transcribed in Chinese characters are underlined in this chapter.

emergence would indicate the commencement of the third stage, a restricted RT system (e.g., 昨天 'yesterday' in 昨天我吃饭了 'Yesterday, I ate.'). Lastly, the study hypothesized that the arrival of the free RT system would be indicated by the children expressing a combination of AMs, TAs, and TNs (e.g., the underlying words in 昨天我吃了饭之后去散步 'Yesterday, I went for a walk after I ate.'), with those three time points relating to each other freely.

After that, to examine how Mandarin-speaking children's acquisition of different temporal markers (form) is affected by content, the acquisition within each category of temporal marker was also examined. It was further hypothesized that different sets of semantic factors would also affect the acquisition of each of the three categories of Mandarin temporal markers. Previous studies have sporadically reported such differences. In light of findings from previous studies, this study proposed that two semantic factors—specificity and remoteness—would affect the acquisition patterns of individual temporal markers within each of the three categories thereof, and that such factors would be useful in explaining acquisition within each of those categories.

As reported in study one, language-sample analysis (LSA) of natural language data was adopted to provide a broad communicative context through freeplay, storytelling using pictures, and conversation with the participants. Specifically, a variety of scenarios replete with opportunities for each child to talk about time points beyond the here-and-now context were created; and all three task types facilitated a maximum observation of language behavior (Rezapour et al., 2011; Southwood & Russell, 2004).

The following research questions were addressed:

1. What are the effects of possessing various time concepts on the acquisition of different categories of Mandarin temporal markers?

2. What are the acquisition trajectories, in terms of both a) emergence and b) mastery, between the three categories of temporal markers among Mandarin-speaking children?

3. What are the effects of a) temporal remoteness and b) specificity on the acquisition within each category of temporal markers?

This current study investigated the acquisition between and within the three temporal-markers categories and examined the effects of time concepts, as well as the semantic features of temporal remoteness and specificity on the acquisition. The findings not only enrich our understanding of the acquisition of temporal markers in children, but also allow the investigation of the content-form interface which contributes to our theoretical understanding of acquisition in the early language of children. Moreover, clinical implications for the intervention for children with language disorders are also implied.

3.4 Methods

3.4.1 Participants

The same participants in study one (chapter two) were involved in the current study, with 82 native Mandarin-speaking typically developing children, aged 25 to 60 months (48 boys and 34 girls), recruited from early education centers and kindergartens in Shenzhen and Guangzhou, China. According to their caregivers, none had any sensory or intellectual disabilities or language problems. The participants were divided into three age groups at one-year intervals (i.e., 25 to 36 months, 37 to 48 months, and 49 to 60 months), and their mean lengths of utterances in words (MLUw) correlated significantly with their ages (r (80) = .336; *p* < .01).

3.4.2 Language Sample Collection and Transcription

Language samples were collected in study one through one-on-one interactions with examiners from three tasks (i.e., freeplay, storytelling using pictures, and conversation)

adopting standardized procedures. They were then transcribed orthographically and analyzed by the trained speech-therapy students and research assistants, all following the same procedures in study one. In particular, deviant utterances were excluded, as children's erroneous use of temporal markers was beyond the scope of this study. The physical context was provided by the descriptions of the events and the participants' actions along with the utterances produced, while the linguistic context was comprised of the examiners' utterances. Both types of contexts were important in ascertaining whether the temporal markers used by the children were semantically correct.

3.4.3 Coding and Classification of Temporal Markers

All utterances coded with *temporal* content following the framework modified from Lahey (1988) were extracted for further analysis, and each temporal marker was classified as an AM, TA, or TN. The members of each of the three categories were then further classified according to their temporal remoteness and specificity. Remarkably, 10% of the language samples were randomly selected and independently coded by a second rater as a check on the inter-rater reliability of the semantic content category coding in study one, which eventually reached a high agreement of 93.0%.

The four subgroups of AMs produced by the participants were perfective *le*, experiential *guo4*, progressive *zai4*, and durative *zhe*. Here, it should be noted that the identification of the perfective *le* was problematic due to its potential overlap with the sentence final particle (SFP) *le* (Liu, 2015). The perfective *le* has a postverbal position and cannot follow a nominal (Zhu, 1982), whereas the SFP *le* is usually located at the end of a sentence (Li & Thompson, 1981). Functionally, perfective *le* asserts both that an event is bounded and that it terminated prior to the time of speech (Liu, 2015) (e.g., 我买了明天的票 'I bought tomorrow's tickets.'). The SFP *le*'s various functions, meanwhile, include marking the reported event or

situation as relevant to the current context (e.g., 他快要上机了 'He's about to board.') and signaling a change of situation/state (e.g., 他当父亲了 'He becomes a father.') (Li & Thompson, 1981; Tang, 2016; Zhu, 1982). Accordingly, this study considered both the function and the position within the utterance when trying to differentiate between the perfective *le* and the SFP *le*. For instance, *le* occurring after a nominal or at the end of a sentence but not in a postverbal position was regarded as the SFP *le*. For ambiguous cases in which *le* occurred both at the end of a sentence and in the postverbal position (e.g., 吃了 'ate'), the conversational context was also taken into account. Inter-rater agreement on temporal-content coding also served as an important check on the consistency of classification. Among the 10,643 major utterances produced by the participants, 1,535 occurrences of *le* were recorded, but only 448 of them were given a final classification as perfective AMs.

As briefly noted above, ambiguity also marks the difference between Chinese TAs and TNs (Biq & Huang, 2016). To differentiate between them, a search was conducted for several unique syntactic properties of TNs that were not found among TAs. First, temporal markers that functioned as arguments in clause structures, subjects, or objects were regarded as TNs (e.g., 今天是我生日; Biq & Huang, 2016; Shi, 2016). Second, a temporal marker was regarded as a TN if it was used immediately after a preposition, such as 在 'at', 到 'to', and 等到 'until' (Yip & Rimmintong, 2016; Zhu, 1982); was modified by a 的 '*de*' phrase (Shi, 2016) (e.g., 星期天 'Sunday' in 开心的星期天 'a happy Sunday'); was modified with a quantity but not with an adverb (Zhu, 1982) (e.g., 一小时 'one hour'); or served as a modifier of another noun/nominal phrase, with or without the use of 的 '*de*' (Shi, 2016) (e.g., 早上 'morning' in 早上的会议 'a morning meeting').

Finally, when dividing each of the three categories of temporal markers into subgroups according to the semantic properties of temporal remoteness and specificity, "temporal

remoteness" was defined as the distance between the ST and the ET (Wagner, 2018); that is, how long ago or how far in the future the events happened or would happen. "Specificity," on the other hand, was defined according to the number of semantic features associated with a given lexical item (Clark, 1973; Pinker, 1989); that is, the more such features a term possessed, the more specific it was deemed to be.

Erbaugh (1992) reported that 96% of the utterances by young children that included perfective *le* were used to mark the immediate past, and Zhou (2004) proposed that the use of *guo4* required the retention of experiences that were more remote from the time of speaking. Therefore, *guo4* should be regarded as more remote than *le*. As for specificity, both perfective *le* and experiential *guo4* mark the termination of events, but using the latter involves an extra specification of a prior experience that was discontinued after a different RT (Liu, 2015). For example, *guo4* in 我去过香港 'I have been to Hong Kong' implies that one was in Hong Kong but is no longer there now, whereas the AM *le* in 我去了香港 'I went to Hong Kong' does not carry that implication. Moreover, in light of the time concepts proposed by Weist (1989), it has been suggested that *le* conveys situations in which ET and RT coincide, in contrast to *guo4*, which conveys that ET precedes RT, which in turn differs from ST (Smith & Erbaugh, 2005). Tang (2016) likewise concluded that *guo4* implies some experience before the RT. As a result, this study considered experiential *guo4* to be more specific than perfective *le* due to the former's additional feature of discontinuation and functional RT concept.

Regarding the imperfective AMs, both progressive *zai4* and durative *zhe* denote continuous/ongoing events semantically, and they are indistinguishable from each other in terms of temporal remoteness. On the other hand, progressive *zai4* is dynamic and denotes events whose final endpoints are knowable, whereas durative *zhe* does not presume an endpoint and is regarded as static. Smith and Erbaugh (2005) also claimed that both progressive *zai4* and durative *zhe* express the same time concepts. In view of the above, both progressive *zai4*

and durative *zhe* in this study were also considered to have equal specificity and were predicted to be acquired at about the same time.

As the ratings of temporal remoteness and specificity for TAs and TNs may vary greatly across individuals, a classification system was developed by asking 10 native Mandarin speakers aged 23 to 40, all of whom held bachelor's degrees, to judge their remoteness and specificity. Temporal markers were assigned to a particular subgroup if 70% agreement was reached among the raters; otherwise, they were deemed not classified and were excluded from analysis (n = 2).

The 23 types of TAs in the dataset were first classified as past, present, and future (Li & Shirai, 2000; Liang et al., 2019). Then, the past and future TAs were each further divided into two subgroups according to whether they were deemed proximate or distant by the same 10 raters mentioned above. For example, the past TA 刚刚 *gang1gang1* 'just' was deemed more proximate than 早就 *zao3jiu4* 'already at an early time', and the future TA 等一会 *deng3yi1hui4* 'wait a moment' was more proximate than 从此 *cong2ci3* 'since then'.

The 22 types of TNs specifying time points and duration were categorized by remoteness and specificity, respectively. For example, time-point TN 晚上 *wan3shang4* 'at night' was deemed more proximate than 星期六 *xing1qi1liu4* 'Saturday', while duration TN 一分钟 *yi1fen1zhong1* 'one minute' was deemed more specific than 一会儿 *yi2huir4* 'a while'.

3.5 Results

Among the 10,643 utterances by the children in the dataset, 679 had temporal tags, collectively containing 850 tokens of temporal markers, and that set of temporal markers comprised 49 different lexical items, including four AMs, 23 TAs, and 22 TNs.

Table 3.1 summarizes the number of temporal markers (tokens and subtypes) produced by the participants. Pearson's correlation testing indicated a significant correlation between age and the token count of temporal markers per utterance (r (80) = .33, p < .001). The number of different subtypes of temporal markers was also significantly higher in the children who were older (r (80) = .41, p < .001), who were also more capable of using more categories of temporal markers (r (80) = .48, p < .001).

Table 3.1 Number of temporal markers (tokens and subtypes) produced, by age group

	Overall (n=82)		Group 1 (n=19)		Group 2 (n=41)		Group 3 (n=22)	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range
Age (months)	43	25-60	30	25-36	42	37-47	55	49-60
Total number of temporal markers	10.38	0-41	6	0-15	10.76	1-31	13.45	3-41
Temporal markers per utterance	.08	0-.32	.04	0-.11	.08	.01-.20	.12	.03-.32
Number of subtypes of temporal markers	4.84	0-17	2.63	0-8	5.05	1-12	6.36	1-17
Number of categories of temporal markers	2.09	0-3	1.42	0-3	2.15	1-3	2.55	1-3

The following subsections will report the findings on the differences in the acquisition of markers in the three temporal categories and on the acquisition of different subgroups within each category, as well as how the subjects used multiple markers within single utterances.

3.5.1 Acquisition Differences between Different Categories of Temporal Markers

Table 3.2 presents the token counts and number of subtypes of temporal markers produced by each of the three sampled age groups.

Table 3.2 Numbers of different categories of temporal markers (tokens and subtypes), by age groups

	Group 1		Group 2		Group 3	
	Mean	SD	Mean	SD	Mean	SD
AMs						
Token	5.05	3.27	6.20	4.66	7.27	4.27
Subtype	1.84	1.07	2.37	1.18	2.64	.90
TAs						
Token	.63	1.46	3.73	4.52	4.05	4.96
Subtype	.53	1.07	2.10	2.14	2.45	2.30
TNs						
Token	.32	.67	.83	1.41	2.14	2.80
Subtype	.26	.45	.59	0.95	1.27	1.49

Note. SD = standard deviation; AMs = aspect markers; TAs = temporal adverbs; TNs = temporal nouns

To investigate the emergence of temporal markers in each category across the sampled children's ages, a one-way analysis of variance (ANOVA) was conducted to compare the number of different temporal-marker types produced across age groups. The results showed a significant effect of age on TAs (F (2,79) = 10.38, $p < .001$) and TNs (F (2,79) = 5.30, $p < .01$). However, the relationship between age and AM type count was non-significant (F (2,79) = 3.21, $p = .07$).

Post-hoc analyses using the Bonferroni test at a significance level of .05 revealed that the three year olds used significantly more examples of TAs than the two year olds did. Likewise, the four year olds produced significantly more examples of TNs than the three year

olds did. Figure 3.1 summarizes the mean occurrence by age group of each category of temporal marker.
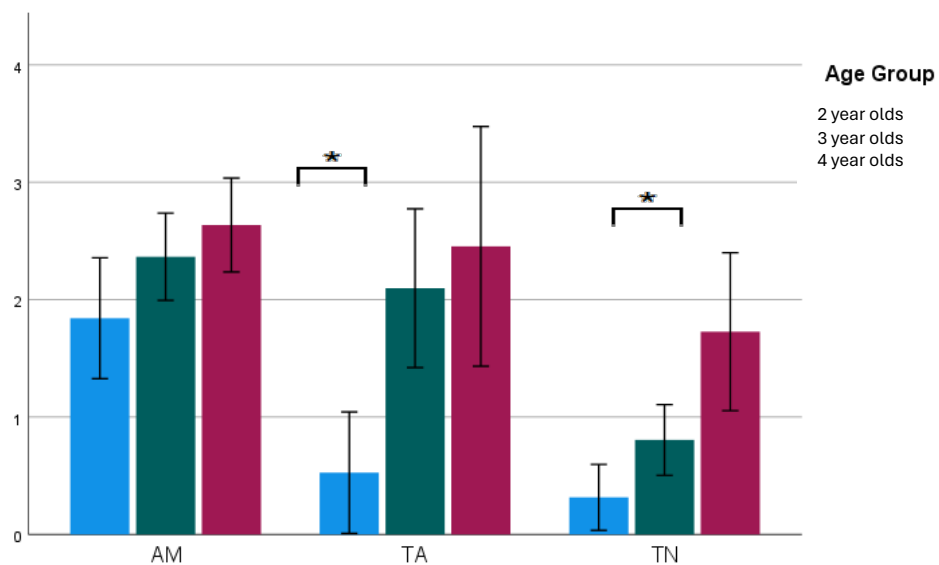
Figure 3.1 Mean frequencies of aspect markers (AM), temporal adverbs (TA) and temporal nouns (TN) across age groups.
Error bars indicate 95% CI, and * indicates $p < .05$ in post-hoc Bonferroni testing

The token counts of temporal markers in each category were then analyzed for evidence of age differences in the mastery of markers. The results of that analysis are shown in Figure 3.2.
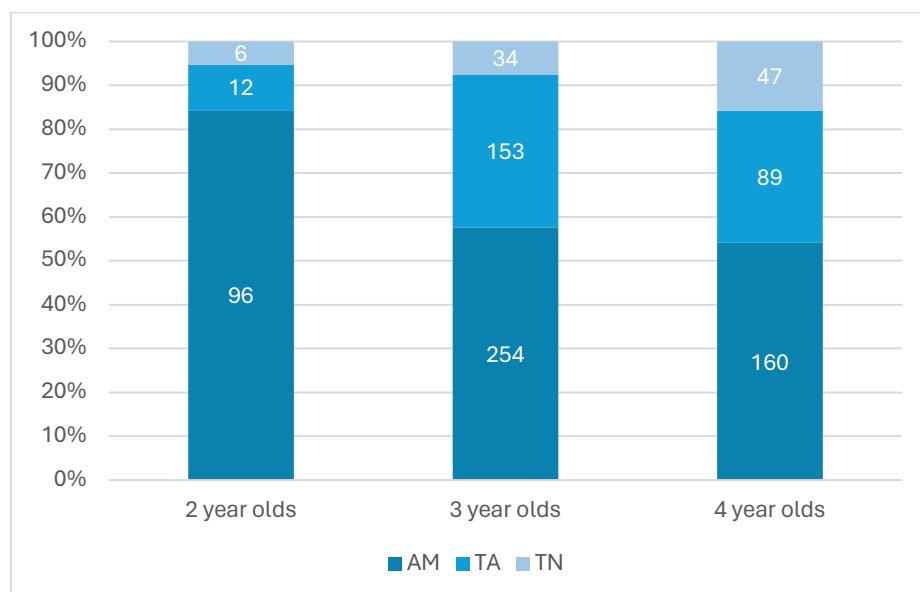
Figure 3.2 Distribution of aspect markers (AM), temporal adverbs (TA) and temporal nouns (TN) produced by each age group

Chi-square testing confirmed that the proportional distributions of the three categories of temporal markers differed significantly across age groups ($\chi^2(4) = 45.78$, $p < .001$). Post-hoc testing was conducted by calculating the differences between the chi-square values to identify which pairwise comparisons contributed to the significant difference (Cox & Key, 1993). To avoid false positive results, Bonferroni correction was adopted: the significance level was divided by the number of tests conducted, yielding .05/6 = .008, in this case. This result indicated that the two year olds used a significantly larger proportion of AMs ($\Delta\chi^2 = 10.81$, $p < .01$) and a smaller proportion of TAs ($\Delta\chi^2 = 10.79$, $p < .01$) than the three year olds did. Moreover, the three year olds used a significantly smaller proportion of TNs than the four year olds did ($\Delta\chi^2 = 7.04$, $p < .008$). However, no significant differences were found between the two year olds' and the three year olds' usage of TNs ($\Delta\chi^2 = .02$, $p = .89$), or between the three year olds' and the four year olds' usage of AMs ($\Delta\chi^2 = 1.30$, $p = .25$) or TAs ($\Delta\chi^2 = 3.47$, $p = .06$).

3.5.2 Acquisition of Different Subgroups within Each Category of Temporal Markers

To examine age-based differences in the production of the subgroups of each of the three temporal categories, point biserial correlation was adopted, which measured the association between a dichotomous and a continuous variable (Kornbrot, 2014). Chi-squared tests and repeated-measures ANOVAs were also used to investigate the distribution of and differences between the number of unique temporal markers in each subgroup.

Aspect markers. The results of the point biserial correlation indicated that the emergence of the progressive marker *zai4* ($r_{pb}(80) = .405$, $p < .001$) and the experiential marker *guo4* ($r_{pb}(80) = .326$, $p = .003$) were significantly more likely in the older children. However, the observed associations between age and the production of perfective *le* ($r_{pb}(80) = .085$, $p = .447$) and durative *zhe* ($r_{pb}(80) = -.116$, $p = .299$) were non-significant.

The proportion of the four AMs produced by each age group are shown in Figure 3.3. Chi-square tests showed that the proportional distributions differed significantly across age groups ($\chi^2(6) = 19.68$, $p < .05$). Pairwise comparisons indicated that the two year olds used a significantly smaller proportion of experiential *guo4* ($\Delta\chi^2 = 5.66$, $p = .017$) than the three year olds did.
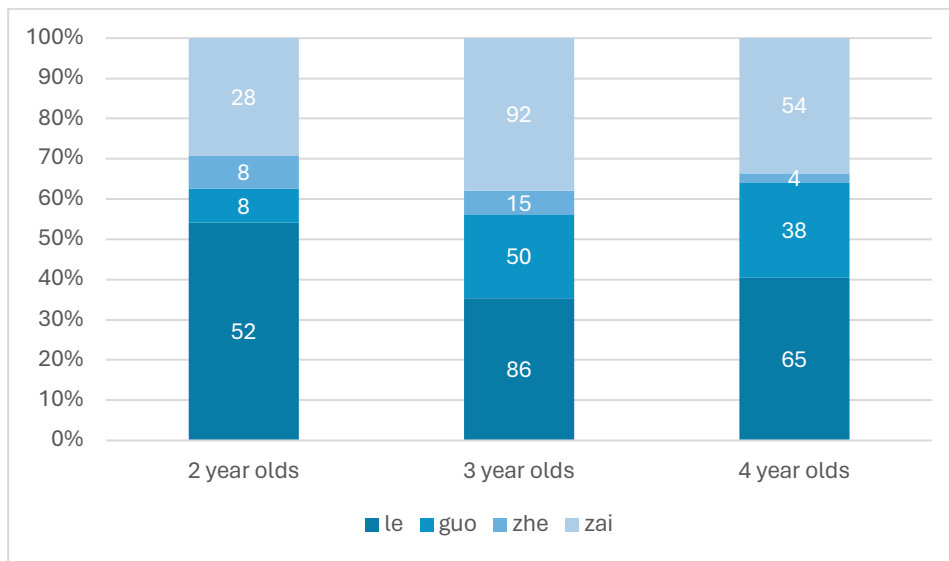
Figure 3.3 Distribution of the four aspect markers across and within age groups

Temporal adverbs. Point biserial correlation revealed that the incidences of distant-past TA use [$r_{pb}(80) = .228$, $p = .039$] and proximate-future TA use [$r_{pb}(80) = .287$, $p = .009$] were both significantly higher in the older children. However, the associations between age and the production of present TAs [$r_{pb}(80) = .131$, $p = .239$], proximate-past TAs [$r_{pb}(80) = .185$ $p = .096$] and distant-future TAs [$r_{pb}(80) = .079$, $p = .481$] were non-significant.

Repeated-measure ANOVAs were performed to investigate the direct and interactive effects of membership of the three age groups, two temporal-distance types, and two timeframes on the number of unique temporal markers in each TA subgroup. The results indicated a significant main effect of age group ($F(2, 79) = 5.57$, $p = .005$): more unique TAs

were observed among the three-year-old children than the two year olds. The main effect of temporal distance [$F(2, 79) = 30.081, p < .001$] was also significant, with more proximate TAs than distant TAs being used.

There was also a significant interaction effect of age group and temporal distance on TA use ($F(2, 79) = 3.467, p = .036$). Post-hoc Bonferroni analysis indicated that among the three year olds, a wider range of proximate TAs were used ($M = 1.41, SD = 1.26$) than distant TAs ($M = .171, SD = .543$), and similar results were also found among the four year olds (proximate: $M = 1.55, SD = 1.90$; distant: $M = .409, SD = .796$) (see Figure 3.4).
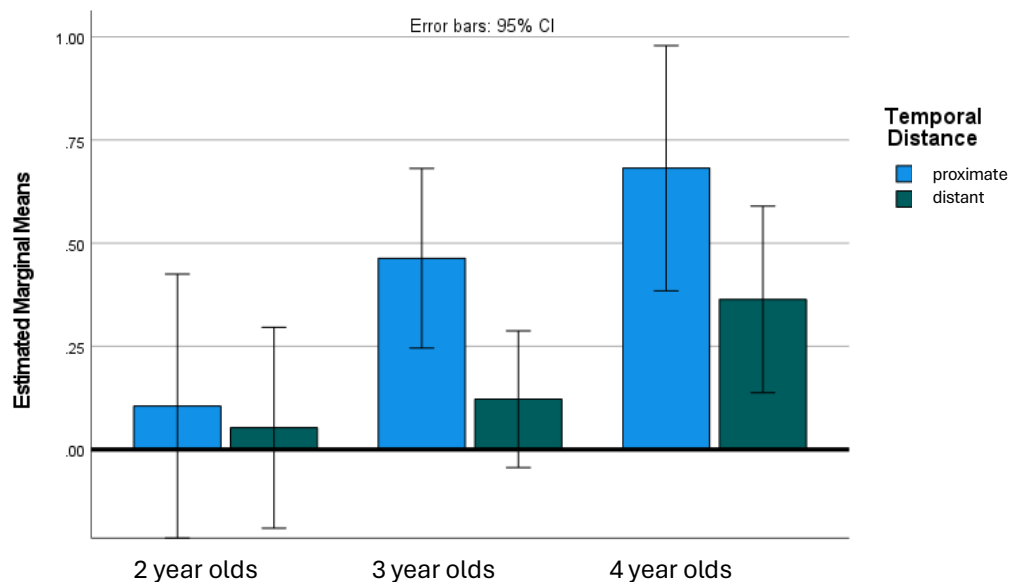
Figure 3.4 Mean number of temporal adverbs in the proximate and distant subgroups, by age group

Finally, the interaction effect of timeframe and temporal distance on the number of unique temporal markers was also significant ($F(2, 79) = 8.863, p = .004$). Post-hoc Bonferroni analysis indicated that the subjects used significantly more different proximate-future TAs ($M = .756, SD = 1.03$) than proximate-past TAs, and more proximate-future TAs than distant-future TAs ($M = .037, SD = .189$) (see Figure 3.5).
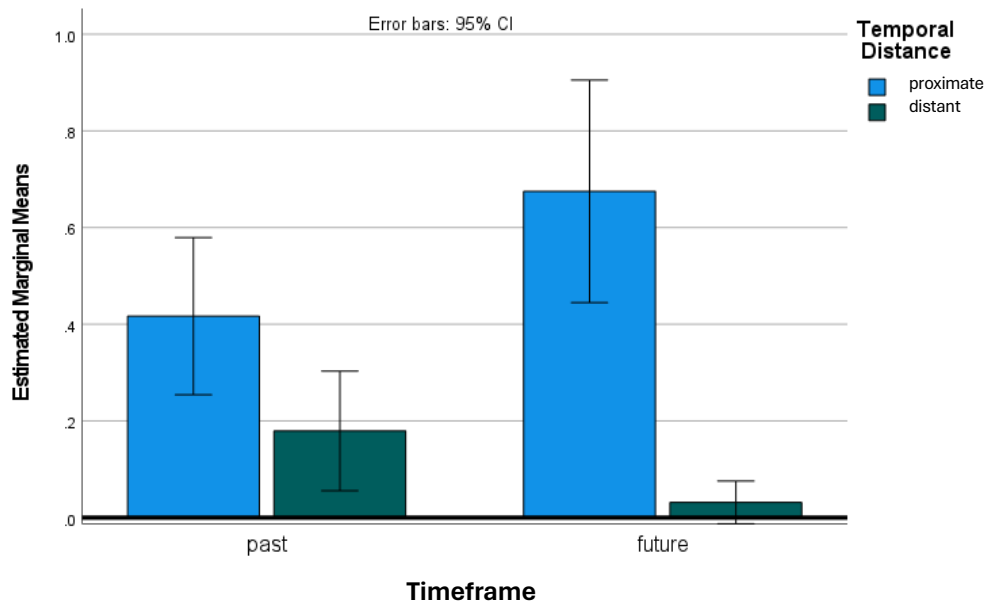
Figure 3.5 Mean number of temporal adverb types in each temporal-distance subgroup

Temporal nouns specifying time points. The point biserial correlation results revealed that significantly more distant TNs ($r_{pb}(80) = .285$, $p = .010$) were used by the older children, but the association between age and the production of proximate TNs ($r_{pb}(80) = .187$, $p = .093$) was non-significant.

A repeated-measures ANOVA was conducted to investigate the direct and interactive effects of age-group membership and the two temporal-distance categories on the number of temporal markers specifying time-points in each TN subgroup. The results indicated a significant main effect of age ($F(2, 79) = 6.27$, $p = .003$), and post-hoc Bonferroni tests further revealed that more TNs specifying time-points were uttered by the four year olds ($M = .864$, $SD = 1.32$) than by the two year olds ($M = .00$, $SD = .00$) and the three year olds ($M = .341$, $SD = .575$) (see Figure 3.6).
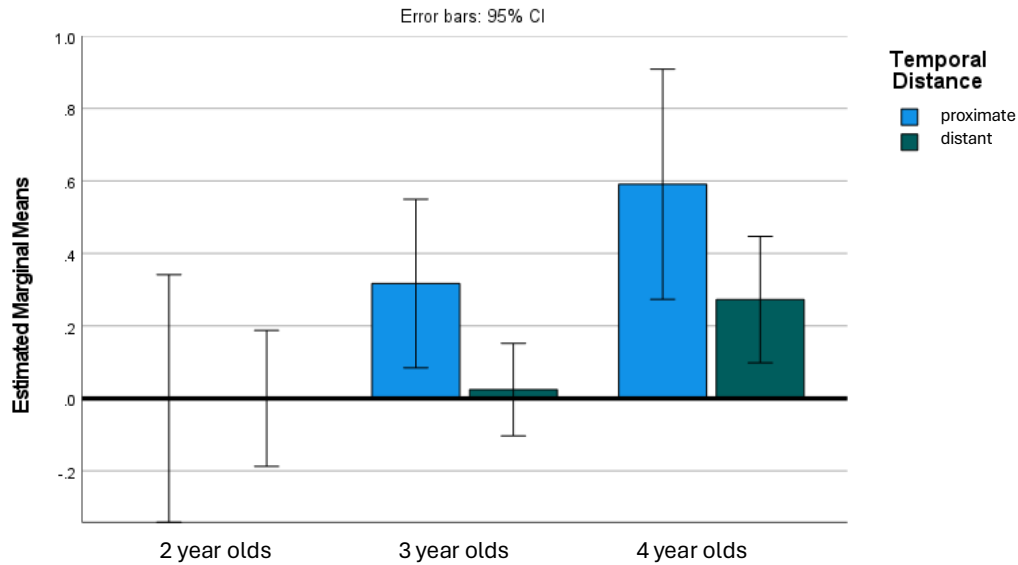
Figure 3.6 Mean number of types of subgroups of temporal nouns specifying time points, by age group

Temporal nouns specifying duration. The point biserial correlation revealed that the emergence of specific TNs ($r_{pb}(80) = .369$, $p < .001$) increased with the children's ages, but here was no significant association between age and the production of generic TNs ($r_{pb}(80) = -.165$, $p = .139$).

A repeated-measures ANOVA was conducted to investigate the effect of age-group membership and specificity on the number of unique temporal markers denoting duration in each TN subgroup, which indicated that there was no significant main effect of either age or specificity (age: $F(2, 79) = 1.33$, $p = .270$; specificity: $F(2, 79) = 1.55$, $p = .217$). However, the interaction effect of age and specificity on the use of TNs denoting duration was significant ($F(2, 79) = 6.90$, $p = .002$). Post-hoc Bonferroni tests further revealed that more specific TNs ($M = .364$, $SD = .492$) than generic TNs ($M = .00$, $SD = .00$) were used by the four year olds. (see Figure 3.7).
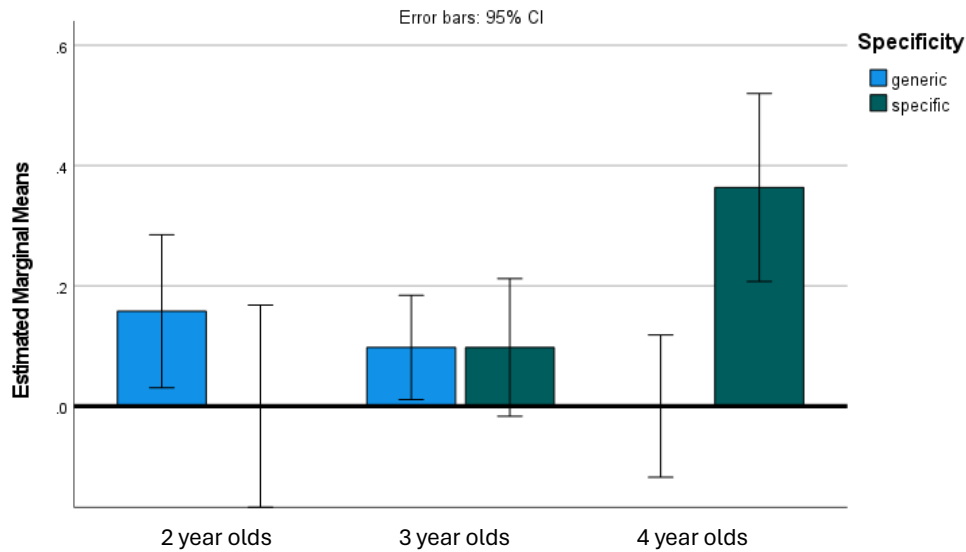
Figure 3.7 Mean number of subgroups of temporal nouns specifying duration, by age group

3.5.3 Use of Multiple Temporal Markers in Single Utterances

Figure 3.8 illustrates the distribution of utterances containing single versus multiple temporal markers. The results of the chi-square testing indicated that these proportions differed significantly across age groups ($\chi^2(2) = 10.91$, $p < .01$). Post-hoc pairwise comparisons (Cox & Key, 1993) with Bonferroni correction (.05/4 = .0125) further showed that the two-year-old children produced a significantly smaller proportion of utterances with multiple temporal markers ($\Delta\chi^2 = 6.77$, $p < .05$) than the three-year-old children did, but that the proportion of utterances with multiple temporal markers did not differ between the three year olds and the four year olds ($\Delta\chi^2$ 2.37, $p = .12$).
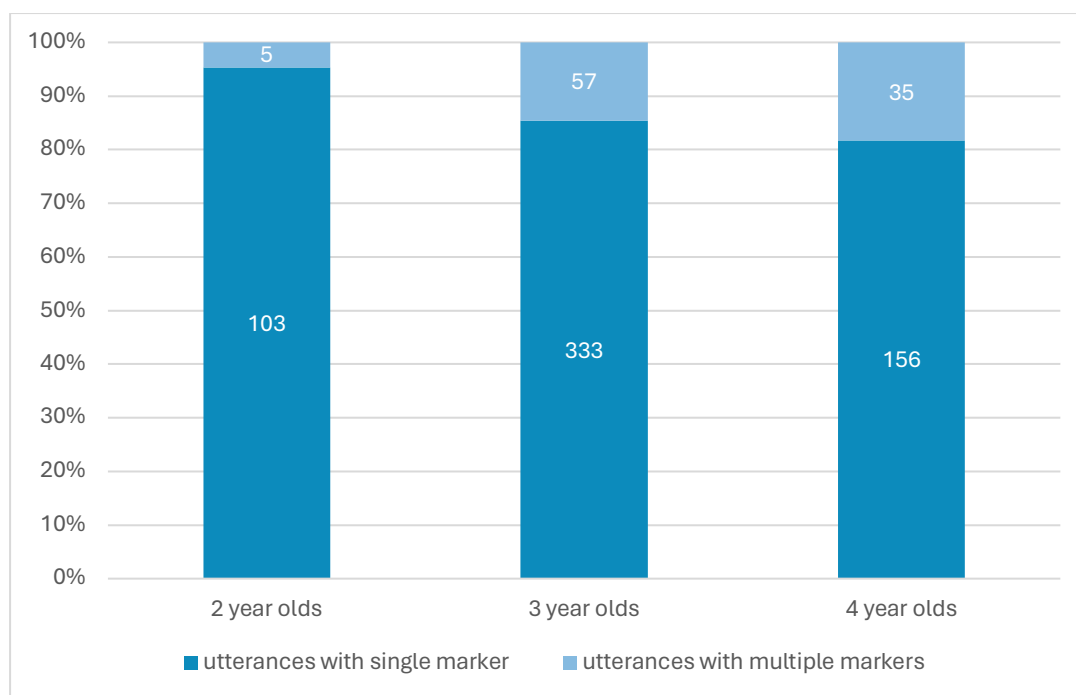
Figure 3.8 Distribution of utterances with single and multiple temporal markers, by age group

The categories of multiple markers used in single utterances were further investigated, and the results are presented in Table 3.3, which shows that the two-year-old children only produced multiple markings by combining markers from two different categories. Both the three and four year olds, in contrast, sometimes produced multiple temporal markings with multiple markers from the same category. Only the four year olds were able to express themselves by combining temporal markers from all three categories.

Table 3.3 Distribution of multiple temporal markings within single utterances, by category and age group

Age group	Single category			Two categories			Three categories
	Multiple AMs	Multiple TAs	Multiple TNs	AM +TA	AM +TN	TA +TN	AM+TA+TM
Group 1	0	0	0	1	2	1	0
Group 2	6	8	1	18	12	9	0
Group 3	4	6	3	9	8	8	1

Note. AMs = aspect markers; TAs = temporal adverbs; TNs = temporal nouns

78

3.6 Discussion

Natural speech production data produced by native Mandarin-speaking preschoolers were used to provide a more comprehensive investigation of acquisition between and within different temporal markers categories in expressing the content of *temporal*. The results indicated three main findings: (1) overall, the advancement in the children's time concepts affected their acquisition of different categories of temporal markers; (2) age-related growth took place in the co-occurrence of the acquisition of the temporal markers in Mandarin-speaking preschoolers; and (3) semantically, more proximate and generic temporal markers tended to be acquired earlier.

3.6.1 Effect of Time Concepts on the Acquisition of Mandarin Temporal Markers

This study's findings upheld its hypotheses on the acquisition of the three categories of temporal markers by the Mandarin-speaking children. According to Weist (1989), various temporal markers are employed when children progress through the four stages of the manipulation of ST, ET, and RT. In the first stage, the ST system emerges between 12 and 18 months, during which children focus on the here-and-now and no tenses or aspects are used. Between 18 and 24 months, children begin to use tenses to mark the ET in the second stage, which is separate from that of ST. Between 30 and 36 months, temporal adverbs are expressed in the third stage, the restricted RT system. Finally, in the fourth stage, the free RT system emerges between 36 and 52 months, and children express ST, ET, and RT freely using a variety of temporal expressions flexibly.

As expected, the findings in this study were consistent with those in the above acquisition trends for different temporal markers in general, as the children developed various time concepts (Weist, 1989). First, the extensive use of AMs by the 2-year-old children was in line with the transition from the first-stage ST system to the second-stage ET system. AMs

were used to denote the internal temporal states of events while the ET concept still coincided with ST in the expressions. Subsequent mastery and a larger repertoire size of TAs in the 3 year olds signaled the commencement of the second-stage ET system. With the expressions of TAs, ET dissociated from the deictic center of ST, resulting in a time displacement between ST and ET. A further mastery and expansion in the repertoire of TNs at age four indicated the anchoring of the RT concept in the third-stage restricted RT system. The time concepts expressed extended from the event itself to a more external timeframe by employing the TNs. A broader time perspective and higher demand in cognitive capacity were involved. Finally, higher co-occurrences of temporal markers in single utterances by older children also supported the beginning of the free RT system. The children employed different types of temporal markers in a single utterance flexibly, to denote a complex timeframe by separating RT from ST and ET. For example, in the utterance 她 刚才 在 喝 'He was drinking a while ago', the progressive *zai4* was used to denote the internal state of the event, while the TN 刚才 *gang1cai2* 'a while ago' was used to anchor a more external RT in the past.

The findings also indicated that only the four-year-olds were able to express temporal markers from all three categories in single utterances, a manipulation and differentiation of ET and RT that clearly required higher-order cognitive ability and capacity. Observations concerning the uses of TNs by the preschoolers were particularly important, as previous studies have seldom mentioned the role of TNs in anchoring RT in the Mandarin temporal system. Likewise, the findings on the usage of multiple temporal markers in single utterances expanded the understanding of how Mandarin-speaking children proceeded from a restricted use of RT to a more independent use. As such, the current study provides empirical evidence that supports how the advancement in children's time concepts affected their acquisition of temporal markers as a whole system.

In addition, the acquisition differences observed among the three temporal-marker categories in Mandarin were also found to be related to their syntactic properties. That is, an AM was expressed with a verb to form a core part of a verb phrase, and it was generally concatenated with a verb or verbal predicate (Liu, 2015; Matthews & Yip, 2011). A more local syntactic structure was involved, as the AM referred to the constituents that made up that structure (Branigan et al., 2006), and such temporal markers were therefore regarded as relatively simple syntactically. On the other hand, TAs sometimes modified a sentence to denote the narrator's temporal perspective (Ernst, 2001), and TNs modified a whole sentence—or multiple sentences—to indicate time points in, or the durations of, the events they described (Yip & Rimmington, 2016; Zhu, 1982). The use of TAs and TNs was comparatively more global than that of AMs, as the former made references to aspects of discourse beyond sentence structure (Branigan et al., 2006). Therefore, both TAs and TNs were regarded as having relatively high syntactic complexity. This explained the later emergence and mastery of TAs and TNs among the children observed in this study.

3.6.2 Acquisition Trajectory among the Three Temporal-Marker Categories in Mandarin

Similar to reports from previous studies (Liang et al., 2019; Tse et al., 2012), the results of Pearson correlation indicated that the acquistion of the three temporal markers by Mandarin-speaking preschoolers is associated with age. That is, the older children were observed not only using temporal markers more frequently, but also in greater variety than their younger counterparts. Importantly, the results further indicate that this pattern also extended to the full range of Mandarin's temporal-marker system. In addition, the older children's expressions included more temporal-marker categories, as well as more frequent usage of multiple temporal markers in single utterances, also echoed the results in the increase in co-occurence of AMs

and TAs in older Mandarin-speaking preschoolers reported by Li et al. (2022), and further extend this to the whole Mandarin temporal system (AM, TA, TN).

Considering the emergence of the temporal markers among the three categories, as indicated by the expanded repertoire in each category, it was found that AMs emerged early among the children, as the results of ANOVA and post hoc analysis revealed that there was no change in the number of different types of AMs produced across age. On the other hand, it was also found that the 2 year olds produced significantly fewer types of TAs than the 3 year olds, and the 3 year olds produced significantly fewer types of TNs than the 4 year olds. That is, the numbers of unique TAs and TNs were still expanding after they had reached age three and age four, respectively.

Regarding the mastery of the temporal system among the three temporal-marker categories, as indicated by the proportional token use of the corresponding temporal-marker category, results of the chi-square test showed that the proportion of AMs uttered by the 2 year olds was significantly higher than that uttered by the 3 year olds. Besides, the 3-year-old children expressed a significantly higher proportion of TAs than the 2-year-old ones. There was also a significantly higher proportion of TNs among the 4 year olds compared with the 3 year olds. The above findings suggested an early mastery of AMs, whereas the mastery of TAs and TNs occurred at later stages of childhood respectively. Previous studies mostly considered the repertoire size of various temporal marker categories produced by young children (e.g., Grant & Suddendorf, 2011; Liang et al., 2019; Tse et al., 2012). Our observations added that the mastery of these three categories of temporal markers also followed a similar acquisition trajectory.

3.6.3 Effect of Temporal Remoteness and Specificity on the Acquisition within Each Category of Temporal Markers

While the development of time concepts and differences in the syntactic properties of different categories of temporal markers explained the overall trend of their acquisition, as reported in previous studies, they were not sufficient to account for the variability that occurred when different items within each temporal-marker category emerged. For example, certain types of TAs (e.g., 从来 *zong2lai2* 'ever') and TNs (e.g., 周末 *zho1umo4* 'weekend') did not emerge until later stages in early childhood. The current study proposed that the seemingly unexpected patterns for terms within each temporal-marker category would be accounted for by the semantic features of temporal remoteness and specificity.

The results confirmed this point; that is, the more proximate and generic temporal markers appeared to be acquired earlier. This was consistent with previous findings (Erbaugh, 1992; Grant & Suddendorf, 2011; Zhou, 2004). Specifically, regarding AMs (see Section 4.2, Aspect markers), the results of the point biserial correlation and chi-squared tests indicated that the experiential *guo4*, which was considered more remote and specific, emerged increasingly with age, and that the older children's time-related utterances contained a significantly larger proportion of *guo4* than the younger children's did. The perfective *le*, on the other hand, emerged early (i.e., at age two), and its production remained relatively stable thereafter. As noted before, the experiential *guo4* is considered more remote and specific, as it denotes events that are more remote from the time of speaking when compared to the perfective *le*, and it also possesses the additional semantic feature of 'termination' and the involvement of a functional reference time concept. As a result, the current findings were consistent with previous ones (Erbaugh, 1992; Grant & Suddendorf, 2011; Zhou, 2004).

Similarly, the results of point biserial correlation revealed a significant positive correlation between the use of progressive *zai4* and age, but an early emergence of the durative

zhe. In this case, however, the semantic features of temporal remoteness and specificity were not applicable to explaining the observed differences, as both *zai4* and *zhe* were indistinguishable in term of the two features. It was speculated that the late emergence of *zai4* may have been related to its preverbal position, which contrasted with the postverbal positions of all the other AMs, and created additional difficulty for the children in acquiring AMs. Future studies should conduct more in-depth investigation of this potential impact of syntactic position on the acquisition of various AMs.

Regarding TAs, the results of the point biserial correlation implied that the use of both distant-past and proximate-future TAs increased significantly with age. The later emergence of distant-past TAs was consistent with the predictions about temporal remoteness. Similarly, the proximate-future subgroup of TAs exhibited increasing use across age groups. The distant-future subgroup was predicted to emerge even later, probably after the age of five, and therefore, that prediction could not be tested using the current study's data.

An interaction effect between age-group membership and temporal distance was identified by the repeated measures ANOVA, with more proximate TAs than distant TAs used by both the three-year-old and four-year-old children. Larger repertoires of proximate TAs in the two older age groups confirmed the expectation that the markers with greater temporal remoteness would be acquired later. Although the children's repertoires of distant TAs were smaller than their repertoires of proximate TAs, it was reasonable to expect that more time was needed to acquire distant TAs; but again, this idea could not be confirmed or disconfirmed in the current study due to the lack of subjects aged 61 months or older in the sample. Future investigations with older participants are therefore warranted.

Although the impact of timeframes has been a common topic in studies of TA acquisition, inconsistent results have been reported (e.g., Bi & Peng, 2002; Kong & Fu, 2004). It is proposed that the inadequacy in the past/future classification framework of TAs caused

disparity in findings about the order of acquisition and therefore a more stringent method that took account of the semantic features of TA subgroups should be adopted. Our findings suggest that temporal remoteness could additionally predict and explain various TAs' acquisition order. Notably, results of repeated measures ANOVA and post hoc analysis indicated an interaction between timeframe and temporal distance, with more types of proximate future TAs than proximate past TAs. suggesting that future ones may emerge earlier in life. It was previously proposed that future TAs are frequently used pragmatically by children to describe their own subsequent actions and intentions (Fu, 2002). This salience of proximate-future TAs may therefore explain their early acquisition, as children's (and indeed, adults') need to express distant-future events is comparatively small. However, future studies that include older participants should seek to confirm this.

Results of point biserial correlation indicated that the occurrence of the time-point TN type, distant TNs increased significantly with age, but that of proximate TNs did not. This can again be explained by temporal remoteness; that is, temporal markers denoting a more distant timeframe emerged at a later age. The results for duration TNs likewise confirmed the prediction concerning specificity; that is, the emergence of specific TNs was significantly and positively correlated with age, but the association between age and the production of generic TNs was non-significant, as revealed by the point biserial correlation. Moreover, results of the repeated measure ANOVA also indicated that the number of specific TNs used by the four year olds was higher than the number of generic TNs they used. It is interesting to note that the two year olds used generic duration TNs only, while the four year olds did not use the generic but only specific duration TNs. This indicated that the acquisition of these specific-duration words emerged at a later stage and gradually replaced the generic ones in denoting time periods, which was also consistent with the claim that children acquired words by continuously adding semantic features to their lexical entries over time (Clark, 1973; Pinker, 1989). In fact, it has

been suggested that the acquisition of specific time words requires formal training in abstract knowledge of clock and calendar time, which children do not usually receive until age six or later (Tillman & Barner, 2015).

Finally, by considering the patterns of different temporal markers' combination into single utterances, it was found that both the three and four year olds produced multiple temporal markings with the multiple use of markers from a single category, whereas the two year olds only produced multiple markings by combining markers from two different categories. It was therefore found that multiple temporal markers from the same category were employed to anchor and specify a particular time point from a semantic point of view. For example, in 我今天下午的课都没了 'My lessons this afternoon were cancelled', two TNs – 今天 *jin1tian1* 'today' and 下午 *xia4wu3* 'afternoon' – were used to specify a more precise timeframe. These findings supported Na's (2017) claim that the co-occurrence of TAs functions to stress and specify the semantic meanings of temporal terms, and then extends it to other categories of temporal markers.

The significant role of temporal remoteness among children's temporal expressions may be explained from the cognitive perspective. According to Trope and Liberman (2003), the storage of distant past events in episodic memory usually contains fewer contextual details. Similarly, the construction of distant future events, which require the recombination of past events into novel scenarios (Addis & Schacter, 2008), should involve the utilization of less contextual details. Addis and Schacter (2008) further proposed that both attention and executive functions are involved in the mental processes, in which decontextualized event requires a higher level of executive function to recombine disparate details into a unique coherent event. Thus, more cognitive resources and effort are required to represent events that are not close to the immediate moment. Young children are expected to be cognitively not mature enough to represent the decontextualized distant events as well as organizing the

episodes into a schema, while the older ones should be more cognitively prepared. As a result, the number of types of temporal markers for recent events was more comparable between the older and younger children, whereas the number of types of temporal markers for distant events was found to be differentiating among different age groups. The demand for the usage of distant TNs and TAs to express decontextualized events among younger kids was thus greatly reduced.

The influence of temporal remoteness among children's time expressions is also consistent with findings from previous neuroimaging studies. Researchers have found that brain regions for storing the past and imaging the future respond differently to event characteristics including temporal remoteness (Addis & Schacter, 2008). In addition, more brain activation in thinking about more temporally distant events is also evidenced (Suddendorf, 2010). These findings of neuroimaging studies reiterated that temporally distant events are more difficult to represent, and confirm the contribution of temporal remoteness in the expression of temporality.

3.7 Concluding Remarks

3.7.1 A New Framework: Content-Form Interaction in the Acquisition of Temporal Markers

The current study examined Mandarin-speaking children's acquisition of temporal markers between and within three categories. It is found that the advancement in children's time concepts affects the acquisition of temporal categories, adding cross-linguistic support to Weist et al.'s (1991) proposal. Besides, age-related growth is also evidenced in the acquisition of Mandarin temporal markers. In addition, temporal markers' semantic features of temporal remoteness and specificity further account for variation in temporal markers' acquisition patterns. The findings constitute an important contribution to the scholarly understanding of patterns of temporal-marker acquisition by young children, both between and within

Mandarin's three categories of temporal markers. The interaction between content and form in the acquisition process is perhaps particularly valuable to consider, though the additional role played by the semantic features of temporal remoteness and specificity also appears vital.

Adopting the content-form framework in Bloom and Lahey (1978), this study hypothesized that time concepts would affect the semantic representation of time and provide the language-general building blocks (content) for the children's use of language-specific linguistic messages (form) to code time. The findings supported that both language-general time concept (content) and language-specific syntactic properties (form) interacted to shape the acquisition of the temporal markers in Mandarin-speaking children in a two-dimensional framework of language acquisition, as illustrated in Figure 3.9. The acquisition between Mandarin temporal marker categories was generally governed by the development of the concepts of time (Weist et al., 1991), and the specific syntactic properties of the Mandarin temporal markers present how different forms of temporal expressions were acquired (AM > TA > TN > multiple markers). Within each temporal-marker category, the children expanded their repertoire of lexical items in an order governed at least partly by the semantic features of temporal remoteness and specificity, such that remote and specific temporal markers were acquired after the proximate and generic temporal markers, respectively.
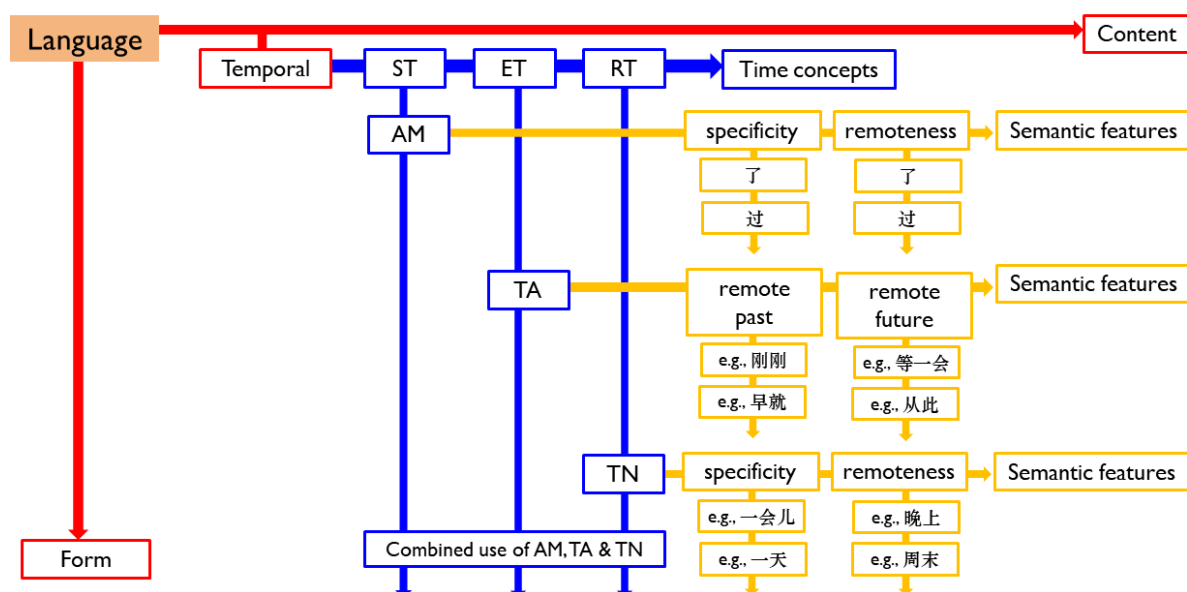
Figure 3.9 An interactive framework of content and form in acquisition

The proposed two-dimensional, content-form interactive framework was supported by the current findings and Weist's (1989) system, and it could also be applied to predict the acquisition patterns of other languages produced by native speakers or language learners; for instance, the acquisition of the (ir)regular tense markings and aspects in English. In addition, the new framework also provided insights that may be beneficial to the language assessment of typically developing children as well as those with language disorders. Time expressions are essential to young children's communication. When assessing such expressions, language content with various time concepts (ST, ET, and RT) and different language forms (AMs, TA, and TNs) should both be examined thoroughly, giving due consideration to the semantic features of temporal remoteness and specificity.

Similarly, the findings of the current study should help to guide interventions for children with language disorders. In particular, a comprehensive and tailor-made intervention plan for temporal expressions that targets syntactically and semantically less complex temporal markers (i.e., AMs) before more complex temporal markers (i.e., TAs and TNs) is suggested;

in addition, more proximate and generic temporal terms should be established before their remote and specific counterparts.

Chapter Four

Manifestations of Language Difficulties at the Content-Form Interface: Case Studies in Mandarin

4.1 Introduction

Children with language disorders exhibit significant difficulties in the acquisition of various aspects of language, which affects their ability to communicate with others effectively. Language disorder is not uncommon in children. It is estimated that around 7-9% of the children, who speak English (Leonard, 2014; Norbury et al., 2016; Tomblin et al., 1997) or Chinese (Wu et al., 2023) are affected. Importantly, language disorders can be found in children as young as toddlers and have detrimental impact on their development. This not only creates problems in achieving different communicative functions, but also causes functional impairment in everyday life including social interaction (Bishop et al., 2017; Leonard, 2014) and learning (McGregor, 2020).

Besides, language disorder is usually associated with poor prognosis (Bishop et al., 2017). It is likely to persist into school-age years and adolescent periods, which significantly affects individuals' academic performance and educational progress. For instance, school-aged children with language disorder often experience difficulties in reading and acquiring basic literacy skills, leading to poor academic attainment (Bishop & Adams, 1990; Catts, 1993; Thompson et al., 2015). Likewise, individuals with language disorder may also be more susceptible to social and behavioral problems (Rice et al., 1991; Yew & O'Kearney, 2013), and even a higher likelihood of involvement in delinquent activities (Brownlie et al., 2004).

Insufficient intervention addressing language disorders can also lead to their persistence into adulthood, resulting in adverse long-term consequences. It can influence individuals'

career prospects and cause unemployment (Law et al., 2009). People with language disorders may encounter difficulties in forming and maintaining relationships throughout their lives as well, leading to a higher chance of experiencing depression and anxiety (Conti-Ramsden &Botting, 2008).

The above creates a strong need for effective intervention for children with language disorders, so that the negative impacts can be minimized. To plan for appropriate intervention, it is essential to have a thorough understanding of the language difficulties present in children with language disorders. Extensive research has shown that there is considerable heterogeneity among children with language disorders (Leonard, 2014; Tomblin et al., 1997). Besides, language disruptions can also manifest in different aspects of language (e.g., Bishop, 1997; Leonard, 2014). Therefore, it would be valuable to investigate the diverse manifestations of language disruptions in children with language disorders. To this end, a significant number of studies were conducted to examine language disruptions in the three basic language domains proposed by Bloom and Lahey (1978), namely Content, Form, and Use. Adopting this content-form framework, the acquisition of semantic content categories, as well as the content-form interface in typically developing Mandarin-speaking children has been reported in study one. Whether these aspects are affected in language disorders remains relatively unknown. Given the above, this study extended the work in study one to examine the language productions of two Mandarin-speaking children with language difficulties. The objective was to investigate if disruptions may manifest in the expression of semantic content categories and the content-form interface. In addition, the study sought to explore the implications of these disruptions and provide insights into effective intervention for children with language disorders.

4.2 Literature Review

4.2.1 Language Disruptions in Children with Language Disorder

As suggested by the literature, language disruptions for children with language disorders can occur in any of the three language domains (i.e., content, form, use), as well as their interactions (Bishop, 1997; Leonard, 2014). This section reviews previous studies that documented disruptions in different domains of child language disorder to indicate how they can be affected in this population. This enhances our understanding of their language disruptions, and eventually informs the planning of appropriate interventions.

4.2.1.1 Studies of Content in Children with Language Disorder

Substantial studies have documented the disruptions in the content domain. In particular, lexical semantics has been identified as a significant area of disruption in children with language disorders. For instance, research has indicated that children with language disorders tend to exhibit lower lexical diversity in their productions and score significantly lower in measures such as the number of different words (NDW), type-token ratio (TTR) and vocD consistently (e.g., Hewitt et al., 2005; Klee, 1992; Klee et al., 2004; Watkins et al., 1995). Besides, it was also found that school-aged children with language disorders demonstrated lower breadth and depth of vocabulary knowledge, in terms of the number of words defined and the amount of semantic information expressed in each correct definition respectively (McGregor et al., 2013; Sheng & McGregor, 2010). Likewise, Oetting et al. (1995) also indicated that children with language disorders demonstrated particular object learning advantage, but difficulty acquiring the action class in quick incidental word learning. Apart from acquiring vocabulary, disruptions in other lexical processing skills were also documented. Brackenbury and Pye (2005) conducted a comprehensive semantic evaluation and

demonstrated that children with language disorders faced big challenges in storing, organizing and accessing lexical knowledge. Some other studies also echoed the impairment in processing and organizing the semantic meanings of words in children with language disorders (e.g., Haebig et al., 2015; Sheng & McGregor, 2010). Similarly, previous studies have also documented disruptions in the specific aspects of the content domain. Alt and Plante (2006) found that preschool children with language disorders encountered problems in fast-mapping different semantic features, while Horvath et al. (2019) also reported difficulties in acquiring certain semantic features of verbs by children with language disorders.

4.2.1.2 Studies of Form in Children with Language Disorder

Considering language form, sentence length has also been studied extensively in language disorders. When assessing sentence length, researchers commonly utilize the mean length of utterances (MLU) as a frequently employed measure. It was found that MLU is developmentally sensitive and acts as an index of normative language acquisition (Rice et al., 2010). On the other side, shorter utterances were found consistently in children with language disorders, in both preschool (Hewitt et al., 2005; Klee, 1992; Rice et al., 2013) and school-age years (McGregor et al., 2015). Besides, the disruptions in syntax and morphology in children with language disorders are also largely documented. Previous studies have indicated that children with language disorders produced utterances with less complex syntactic structures than typically developing children (Mainela-Arnold & Evans, 2014; McGregor et al., 2015). They may struggle to understand and produce complex sentence structures, such as embedding and coordination (Rice & Wexler, 1996). Similarly, lower accuracy in producing various morphological constructions including verb inflections (Rice & Wexler, 1996), past tense morphology (Rice et al., 1998), plural and possessive forms (Kamhi, 2014), and the use of

derivational morphology (Marshall & Van der Lely, 2007) were also found in children with language disorder.

4.2.1.3 Studies of Use in Children with Language Disorder

Finally, disruptions in language use in children with language disorders have also been proposed. Cummings (2009) suggested that children with Autism Spectrum Disorder may experience disruptions in language use, which are secondary to the challenges in linguistic form and content. These children may have limited receptive and expressive language skills, and struggle to comprehend and utilize conventional means of production for specific speech acts. On the other hand, primary disruptions in language use in children with language disorders have also been evidenced. Bishop et al. (2000) found that children with specific language impairments exhibit difficulties in maintaining smooth and coherent conversations with adults. The study highlighted that children with language disorders often exhibited delayed responses, difficulties in initiating and sustaining conversational turns, and challenges in providing relevant and meaningful contributions to the ongoing discourse. These conversational difficulties were not solely explained by the limitations in grammar and vocabulary, indicating that there are additional underlying factors that contribute to their impaired conversational responsiveness. Besides, Rinaldi (2000) also claimed that disruptions of language use can occur independently in children with relatively intact semantic language skills. Specifically, ten students with language disorders were found to have difficulty using context to understand implied meanings, despite their age-appropriate performances reflected by standardized vocabulary assessment.

4.2.1.4 Studies of Mandarin-Speaking Children with Language Disorder

As in the context of Mandarin, a growing body of literature has investigated the disruptions in various language aspects in Mandarin-speaking children with language disorders. Consistent with findings with English-speaking children, Wu et al. (2019) reported lower lexical diversity in the language samples of Mandarin-speaking preschool children with language difficulties. In the realm of lexical semantics, Hao et al. (2008) and Tardif et al. (2009) examined the acquisition of different semantic categories in the early vocabulary inventory of young Mandarin-speaking children using parent report questionnaires. Ma et al. (2009) on the other hand investigated the effect of the semantic measure of imageability on the acquisition of verbs, revealing that imageability significantly predicted the age of verb acquisition in Mandarin-speaking children.

Likewise, studies have also replicated the findings on shorter sentence length among Chinese-speaking children with language disorders (e.g., Klee, 1992; Wu, 2020). Besides, poorer performance in various morphological and syntactic constructions like aspect markers, classifiers (Cheung, 2009), and passive sentences (Durrleman et al., 2023) was also found in Chinese children with language disorders. Huang et al. (2022) investigated early grammatical markings in Mandarin-speaking toddlers. It was observed that more perceptually salient and obligatory markers were acquired earlier. More importantly, it was found that the acquisition trajectories of these markers differed from those in English and other languages, highlighting the linguistic specificity in the acquisition of morphology. Hao et al. (2018) investigated the production of narratives by Mandarin-speaking children with and without language impairment. Results showed that those with language impairment exhibited poorer abilities in syntactic complexity and lexical diversity than their typically developing peers.

All the above findings indicate a diverse manifestation of disruptions across different domains of Mandarin-speaking children with language disorders. Given the heterogeneity observed among children with language disorders, it is crucial to comprehensively profile their language disruptions. This contributes to improving our knowledge of language disorders, and ultimately aids in the selection of appropriate intervention goals.

4.2.2 Potential Disruptions in Semantic Content Category in Children with Language Difficulties

As previously discussed, children with language disorders often experience disruptions across various linguistic domains, including the content domain. Previous studies documenting language difficulties within the content domain have predominantly focused on lexical semantics (e.g., Alt & Plante, 2006; McGregor et al., 2013). These studies have made substantial contributions to our knowledge of language disorders and go beyond the disruptions in surface form. To deepen our understanding of language disorders, it is crucial to explore if the disruptions extend beyond the lexical level of the content domain. This examination further enriches our knowledge of language disorders and inspires an effective intervention.

Considering the potential disruption in expressing semantic content beyond lexical level by children with language disorders, studies in this area are quite rare. Among the few existing studies, Stockman (1996) investigated the production of a child with language difficulty, on a basic set of semantic content categories proposed in Brown's (1973) cross-cultural studies. These semantic content categories represent a restricted set of common topics and ideas in children's early expressions, regardless of their native language (Bowerman, 1973; Brown, 1973). Results showed that fewer types of semantic content categories were expressed in general when compared with typically developing children. In addition, the productivity of

the semantic content category was also investigated in that study by considering the number of exemplars and situations in which the semantic content category was being expressed. It was found that fewer types of semantic content categories were produced productively by the child with language difficulty than by the children with typical development. These findings indicated the potential manifestations of language disruptions in expressing semantic content category in children with language disorders.

In study one, the acquisition trajectory of various semantic content categories produced by Mandarin-speaking children with typical development was investigated. By adopting a 90% acquisition criterion, some semantic content categories were found to be acquired earlier, while others appeared later in early childhood. Although disruptions in the expressions of semantic content categories have been documented in English-speaking children with language difficulties (Stockman, 1996), to our knowledge, no studies were conducted to investigate the potential disruptions in Mandarin-speaking children with language disorders.

Considering the heterogeneity that exists in language disorders, it is speculated that there are Mandarin-speaking children with language disorders who may demonstrate difficulties in expressing certain semantic content categories. To enhance our understanding of language disorders and explore the diverse manifestations of language disruptions, this study aims to investigate if disruptions may manifest in the expression of semantic content categories in Mandarin-speaking children with language difficulties. This investigation will refer to the findings on the acquisition of semantic content categories by typically developing Mandarin-speaking children reported in study one.

4.2.3 Potential Disruptions in the Content-Form Interface in Children with Language Difficulties

In child language acquisition, not only do content and form alone play an important role, some studies have also proposed that semantic content and language form interact in the acquisition process. The content-form interface refers to the relationship and integration between knowledge of the semantic meaning and the linguistic form of language. Children acquire the two domains together and this content-form interface provides the basis of their early expressions (e.g., Bloom, 1991; Bowerman, 1973). Nevertheless, the significance of the content-form interface in language disorder received relatively less attention compared to the separate examination of each domain individually. In the study by Mok and Kipka (2009), the important role of content-form interface with a 5;3 girl diagnosed with communication disorder was illustrated. By analyzing the language samples obtained, the child was found to produce various syntactic errors. These "apparent" syntactic problems, however, were proposed to be attributed to breakdowns in the content-form interface. In other words, the child demonstrated disruptions in the integration of knowledge on both content and form. Accordingly, the failure to use appropriate syntactic structures for the lexical items was caused by the immature semantic representations (content) being carried over to the lexical representations of the words (form). For example, the verb 'know' has different meanings and each meaning is associated with various types of compliments. The ability to produce different complement structures thus reflects the mastery of the semantic meanings of the verb. In contrast, a limited range of compliment structures found in the subject may indicate an incomplete knowledge of the verb's meaning (Pinker, 1989). It further suggested that the surface syntactic errors may probably reveal the disruption in content that was manifested through the content-form interface. Mok and Kipka (2009) urged that this content-form interface indeed plays an important role in child language disorders. Thus, the disruptions in the content-form interface should be taken into

account during the evaluation of children with language disorders, which may otherwise be covered up as syntactic errors and eventually be overlooked during intervention planning.

Study two also demonstrated the content-form interface in language acquisition. This study investigated the expression of a particular semantic content category – *temporal* – by typically developing Mandarin-speaking children. Results indicated that the acquisition among the three categories of temporal markers, namely aspect marker, temporal adverb and temporal noun, was generally governed by the development of time concepts (Weist, 1989). These time concepts play a crucial role in the semantic representation of time and provide the building blocks (content) that children use to express time through linguistic symbols (form). In addition, the acquisition of different lexical items within each category of temporal marker also varied according to the semantic features of temporal remoteness and specificity. Taking considerations in the acquisition of both language content and form, a two-dimensional framework of language acquisition was proposed, which emphasizes the content-form interface in early language acquisition. In another study, Negen & Sarnecka (2012) investigated the acquisition of numbers by preschool children. It was proposed that after scaffolding the number concept through visuospatial development, the content of the number was integrated with forms to create the number words. Importantly, this number word acquisition was supported by the acquisition of general vocabulary, as children's noun knowledge helped them to identify the referents of number words. The acquisition of number words could therefore be supported by being able to pick up both the semantic and syntactic cues from expressing the nouns.

Despite the above-mentioned important role of the content-form interface in child language, apparently no studies were conducted to investigate its manifestations in Mandarin-speaking children with language disorders. Extensive research has consistently shown that children with language difficulties often demonstrate noticeable disruptions of form. These disruptions typically manifest as shorter utterances (e.g., Hewitt et al., 2005; McGregor et al.,

2015) and a limited vocabulary repertoire (Klee et al., 2004; Watkins et al., 1995). In view of the above, it is speculated that these form disruptions may lead to difficulties in expressing certain semantic content categories and/or difficulties in acquiring certain forms to express different semantic content categories. The current study therefore aimed to investigate if there are Mandarin-speaking children with language difficulties who may demonstrate disruptions in expressing certain semantic content categories that lie in the content-form interface.

4.3 The Current Study

Diverse manifestations of disruptions have been found within and across various linguistic domains in children with language disorders. It is of paramount importance to profile their language abilities comprehensively so that appropriate intervention can be applied to minimize the negative impact on the overall development. While the majority of previous studies on child language disorder had been focused on the study of form or lexical semantics (e.g., Kamhi, 2014; McGregor et al., 2015), relatively few studies examine the semantic domain beyond the lexical level. In particular, there is currently no study investigating the production of semantic content categories among Mandarin-speaking children with language disorders. How the productions of these semantic content categories are affected in Mandarin-speaking children with language disorders is relatively unknown. On the other hand, the disruptions in content-form interface in children with language disorder suggested by Mok and Kipka (2009) are also relatively unexplored. Therefore, the current study aims to examine if disruptions in the expressions of semantic content categories, as well as in the content-form interface, may occur in Mandarin-speaking children with language difficulties.

Before achieving the goal of investigating the expression of semantic content categories in Mandarin-speaking children with language disorders, the two participants were tested on

traditional measures of sentence length and lexical diversity, to determine if the two studied cases referred by parents and a local therapist, were potentially having language disorder. It is predicted that both participants exhibit shorter utterance length and lower lexical diversity in general.

Next, considering the acquisition trajectory for various semantic content categories among typically developing Mandarin-speaking children reported in study one, the presence of age-expected semantic content categories among Mandarin-speaking children with language disorders was examined to explore the potential disruption. Likewise, extending the work of study two on the acquisition of content-form interface to language disorder, the content-form interface was examined by investigating the diversity of semantic content categories expressed by a particular language form (verb), as well as the diversity of lexical items and syntactic structures expressed in the specific semantic content category (*dative*) in Mandarin-speaking children with language difficulties. The current study explored the following:

1. Are there any Mandarin-speaking children with language difficulties who demonstrate impairments in exhibiting certain age-expected semantic content categories?

2. Do Mandarin-speaking children with language difficulties demonstrate impairments in the content-form interface, in terms of

    a. using a particular form to express a diversity of semantic content categories?

    b. expressing a diversity of lexical items in the particular semantic content category?

The novelty of the current study is to provide empirical data on the expression of semantic content categories by Mandarin-speaking children with language difficulties, and to enrich their language profiles from the semantic perspective. By comparing their production of semantic content categories with those expressed by children with typical development,

potential manifestations of disruptions in the content domain were explored. In addition, our results also shed new light on the content-form interface in the disordered language, which importantly guides the intervention planning.

4.4 Method

4.4.1 Participants

Two native Mandarin-speaking children with language difficulties participated in the current study. Both of them were referred by a local therapist in Guangzhou, China as typical cases with language difficulties. They were reported to have limited language output to express their needs in kindergartens, which affects their academic performance and social interactions with other kids according to their caregivers, with the absence of any reported sensory or intellectual disabilities.

The first participant, LY, is a 3;11 girl studying K1 in a local kindergarten in Guangzhou. According to her parents, LY understood the others well but mainly expressed herself with single words or word combinations. In addition, she also demonstrated limited vocabulary in her expressions. Another participant, PH, is a 4;02 boy studying K2 in another local kindergarten in Guangzhou. PH's parents reported that the child can follow others' commands, but only expressed with short phrases in daily communication. On some occasions, he tended to have difficulties expressing himself, especially when he was asked to describe what had happened in school. Both are receiving language treatment in a local speech therapy training center.

The participants' nonverbal intelligence was evaluated by the Primary Test of Nonverbal Intelligence (PTONI, Ehrler & McGhee, 2008), which is utilized to evaluate the

cognitive ability of children aged 3 to 9 years old. LY got a standard score of 126 and performed at the 96th percentile. On the other hand, PH got a standard score of 102, which is at the 55th percentile. According to Leonard (2014), it is difficult to identify language disorders in young children due to great heterogeneity. Given the relatively intact cognitive abilities and significant functional challenges they experienced, both LY and PH were identified as individuals with language difficulties (LD), potentially indicating the presence of a language disorder (Bishop et al., 2017).

4.4.2 Language Sample Collection and Transcription

The two LD participants' language samples were collected in a quiet room by a local native Mandarin-speaking therapist, who is familiar with them and had received prior training on the data collection procedures. The procedures of language sample collection were identical to those reported in study one, with a warm-up period followed by the three tasks of taking language samples (i.e., freeplay with toys, storytelling with pictures, and conversation) using a standardized protocol. All sessions were audio- and video-recorded and the orthographic transcription of all utterances produced by the LD participants and interviewer was done by a speech therapist, following the same procedures in study one. Physical context was provided by the descriptions of events and the participants' actions along with the utterances produced, while linguistic context comprised the examiner's utterances. In addition, the parts of speech were annotated in each word of all child utterances and the assignment of semantic content categories (Lahey, 1988) in each child utterance was also done by the same speech therapist.

4.4.3 Data Analysis

Each participant's major utterances were used to calculate the mean length of utterance in word (MLU), following Cheung's (1998) and Zhu's (1982) procedures. Traditional measures on lexical diversity of the LD participants including the number of different words produced (NDW), the total number of words produced (TNW), type-token ratio (TTR), vocD and the number of different open/closed class words were also calculated. The production of different types of semantic content categories by the LD participants was also reported. In addition, to investigate the content-form interface, comparisons of different semantic content categories expressed by verbs, as well as different lexical forms and structures used to express the semantic content category of *temporal* between the typically developing participants (TD) reported in study one and the two LD participants were also conducted.

4.5 Results

4.5.1 Utterance Length and Lexical Diversity

Table 4.1 presents the results of measures of MLU, NDW, TTR and vocD of the two LD participants. In comparison with similar measures reported in study one obtained from children with typical development (TD), the results indicated that LY performed one standard deviation below the average achieved by the 3-year-old group in the measures of MLU [mean = 3.61, S.D. = .82, as reported in study one], NDW [mean = 168.26, S.D. = 40.45, as reported in study one], and vocD [mean = 41.10, S.D. = 12.0, as reported in study one]. Similarly, PH also performed lower than one standard deviation below the average achieved by the 4-year-old group in the measures of MLU [mean = 3.78, S.D. = .77, reported in study one], NDW [mean = 163.05, S.D. = 45.26, reported in study one], and vocD [mean = 48.30, S.D. = 6.93, reported in study one].

Table 4.1 Different language measures of the participants with language difficulties

	LY	PH
Mean Length of Utterance	1.98	2.19
Number of Different Words	42	87
Total Number of Words	105	262
Type Token Ratio	0.40	0.33
Total Number of Utterances	46	97
vocD	19.75	37.71

Table 4.2 shows the number of different open class words (nouns, verbs and adjectives) and closed class words produced by the LD participants. In comparisons with similar measures reported in study one obtained from TD children, the results indicated that LY's performance fell below one standard deviation below the average achieved by the 3-year-old group peers in terms of the number of different nouns produced [mean = 47.4, S.D. = 12.67, as reported in study one], number of different verbs produced [mean = 34.05, S.D. = 9.86, as reported in study one], number of different adjectives produced [mean = 17.19, S.D. = 6.33, as reported in study one], and number of different closed class words produced [mean = 69.52, S.D. = 16.02, as reported in study one]. Similarly, PH also performed worse than one standard deviation below the average achieved by the 4-year-old group peers in terms of number of different verbs produced [mean = 32.36, S.D. = 9.51, as reported in study one], number of different adjectives produced [mean = 17.95, S.D. = 5.71, as reported in study one], and number of different closed class words [mean = 64.95, S.D. = 18.68, as reported in study one].

Table 4.2 The number of different lexical items produced by participants with language difficulties

	LY	PH
Number of different noun	18	49
Number of different verb	12	15
Number of different adjective	6	7
Number of different closed class words	18	24

4.5.2 Expression of Various Semantic Content Categories

In study one, a 90% criterion was adopted to establish the age of acquisition for different semantic content categories by the TD children. Specifically, a semantic content category was regarded as acquired by the particular age group if 90% of the participants in the group produced the semantic content category at least once in the sample collected. The expression of various semantic content categories by the two LD children was also recorded and compared with the reported acquisition trajectory. Table 4.3 shows the age ranges of acquisition of various semantic content categories by the TD children and the production by the two LD participants. It was observed that LY demonstrated a comparable set of semantic content categories with her age-matched peers, while certain semantic content categories (i.e., *denial, additive* and *causal*), which were expected in the corresponding age ranges, were absent in PH.

Table 4.3 Semantic content categories produced by LY and PH and the corresponding expected age range of acquisition.

| Semantic content category | Participants with language difficulties (age) | | Expected age range of acquisition (months)^ |
	LY (3;11)	PH (4;02)	
Existence	*	*	24 - 36
Non-existence	*	*	
Reject	*	*	
Attribute	*	*	
Action	*	*	
Locative state	*	*	
State	*	*	
Quantity	*	*	
Temporal	*	*	
Denial	*	○	37 - 48
Possessive	*	*	
Locative action	*	*	
Dative	*	*	49 - 60
Additive		○	
Causal		○	
Notice			Beyond 60
Recurrence			
Specification			
Adversative			
Epistemic			
Communication			

^ Expected age range as reported in study one

* present; ○ expected but absent

4.5.3 Expressions in the Content-Form Interface

Expressions in the content-form interface were investigated by analyzing the variety of semantic content categories expressed using verbs and the diversity of lexical forms in expressing the semantic content category of *temporal* by the two participants. Details of the results are shown as follows:

4.5.3.1 Expressing Different Semantic Content Categories with Verb

Various semantic content categories represented by the verbs of LY and PH were examined, to investigate the disruption in content-form interface. Study one has suggested that typically developing children used verbs to represent seven different semantic content categories, including *action, state, locative action, dative, notice, communication* and *epistemic*. Table 4.4 presents the number of different lexical items produced in these semantic content categories expressed through verbs in the LD participants. Among the seven semantic content categories expressed by verb, three of them were expected in LY's age but she only expressed *action* and *locative action* using verbs, while *state* was absent. In addition, only two different types of verbs were used to express the content of *locative action*. Similarly, PH was able to use verbs to express all the age-expected semantic content categories (i.e., *action, state, locative action, dative*). However, only one verb type was used to express *dative* in his entire language sample.

Table 4.4 Number of different lexical items produced in various semantic content categories expressed by verb in the participants with language difficulties.

Semantic content category	Participants with language disorder (age)		Age range of acquisition (months)^
	LY (3;11)	PH (4;02)	
Action	15	41	24 - 36
State	0	3	
locative action	2	8	37 - 48
Dative	1	1	49 - 60
Notice	0	0	Beyond 60
Communication	0	0	
Epistemic	0	0	

^ Expected age range as reported in study one

4.5.3.2 Expressing Semantic Content Category of Temporal with Different Forms

The linguistic forms employed to express the *temporal* content by children with language difficulties were also examined in the current study. Table 4.5 indicates the number of different lexical items produced in each temporal marker group by the LD participants. It was found that both LY and PH were only able to employ aspect markers to express *temporal* content. The use of neither temporal adverb nor temporal noun was recorded. Also, the types of aspect markers being employed were relatively limited, with only two types being used by LY (perfective 了 *le* and durative 著 *zhe*) and one type being used by PH (perfective 了 *le*).

Table 4.5 Number of different lexical items produced in different temporal marker categories in expressing *temporal* by the participants with language difficulties.

Temporal marker group	Participants with language difficulties (age)	
	LY (3;11)	PH (4;02)
Aspect marker	2	1
Temporal adverb	0	0
Temporal noun	0	0

4.6 Discussion

In the current study, the language profiles of two Mandarin-speaking children with language difficulties analyzed in terms of content and form produced were reported. In comparisons with the performances of TD children of similar age ranges reported in study one, results regarding their manifestations in shorter utterance length and lower lexical diversity tended to indicate that the two cases referred by parents and a local therapist are potentially having language disorder (e.g., Hewitt et al., 2005; Rice et al., 2013). Besides, results regarding the expressions of various semantic content categories also contributed additional understanding of the manifestations of disruptions in the content domain and the content-form interface. Details were discussed in the following.

4.6.1 Utterance Length and Lexical Diversity

The current study showed that the two children with language difficulties produced relatively shorter utterances and lower lexical diversity than most of their age-matched peers with typical development, as reflected by MLU, NDW and vocD. Besides, both participants

also produced fewer different types of verbs, adjectives and closed class words than most of the age-matched peers.

Previous literature has suggested the clinical importance of utterance length and lexical diversity in language disorders (e.g., Hewitt et al. 2005; Klee, 1992; Rice et al., 2010; Watkins et al., 1995). According to Owen and Leonard (2002), a measure should be sensitive to developmental differences to be capable of reflecting clinical differences between children with typical development and those with language disorders. In study one, it was observed that higher MLU, NDW, TTR and vocD were associated with older children in general, which suggests that utterance length and lexical diversity were able to demonstrate developmental differences in the typically developing Mandarin-speaking children. On the other hand, findings on shorter utterance length and lower lexical diversity in the two LD participants in the current study are in line with the findings in the literature (e.g., Hewitt et al. 2005; Klee, 1992). As mentioned before, parents had reported significant impacts and slow progress on the educational performances and social interactions of the children, which may be considered "poor prognostic indicators" in language disorders. Taking into consideration both functional impairment in everyday life and poor prognosis, the findings on utterance length and lexical diversity further indicate the possibility for the two participants to exhibit language disorder (Bishop et al., 2017).

It is noteworthy that both participants with language difficulties demonstrated lower NDW and vocD but not lower TTR in comparison with the age-matched typically developing peers. It is speculated that the small number of utterances produced by children with language difficulties may account for the comparable TTR. In particular, less than fifty utterances were recorded in one of the LD participants, LY. With fewer utterances being elicited, the diversity of lexical items may appear larger with higher TTR. In fact, some researchers have argued that TTR is not sensitive enough to differentiate children with and without language disorders (Klee,

1992; Watkins et al., 1995). Therefore, the current results tended to support the claim that NDW and vocD are preferable and appear to be more appropriate diagnostic indicators of lexical diversity in disorder research (Klee, 1992; Klee et al., 2004; Watkins et al., 1995).

Apart from the two traditional measures of utterance length and lexical diversity, a more comprehensive language profile should be analyzed for each individual child, so that appropriate treatment can be planned accordingly. One possible way to conduct language profiling involves further categorization of lexical items produced by the children, for example, into open class and closed class words. Open class words primarily convey the concrete content of the sentences whereas closed class words are usually more related to the grammatical aspects of sentences, and include relatively few members (Weber-Fox & Neville, 2001). In the current study, it was shown that both LD participants produced fewer types of verbs, adjectives and closed class words than most of the age-matched peers. These findings, especially for the production of closed class words that often serve grammatical functions, were also consistent with previous findings (e.g., Grela & Soares, 2004; Stokes & Fletcher, 2000). To this end, apart from looking into the different types of vocabulary items produced by the children, the analyses of semantic content categories produced also provide another way of conducting language profiling. Details are elaborated as follow.

4.6.2 Manifestations of Disruptions in Expressing Semantic Content Category

Shorter utterance length and lower lexical diversity in children with language disorder are relatively easily detected (e.g., Hewitt et al. 2005; Klee et al., 2004; Watkins et al., 1995). Apart from these measures, investigating the expressions of various semantic content categories by children with language disorders could supplement our understanding of various manifestations of language disruptions. By comparing the production of semantic content

113

categories in children with language disorders to the acquisition trajectory obtained from typically developing children in study one, it was found that interestingly, semantic content categories expected from the language produced by children of the corresponding age ranges, were all present in only one, instead of both, of the LD participants.

LY, despite having shorter utterances and lower lexical diversity, exhibited no difficulties in expressing semantic content categories compared with age-matched peers. On the other hand, apart from demonstrating shorter utterance length and lower lexical diversity, PH's language problems appeared to be also significant in terms of the variety of semantic content categories that were produced. Specifically, the content of *denial* was expected to be acquired in the 3 year olds, but found to be absent in PH, who was four-year-and-two-month old. Likewise, the content of *additive* and *causal,* which were expected in the 4 year olds, were also missing in the language samples of PH. This observation was consistent with the findings of a previous study which reported fewer types of semantic content categories produced by a child with language disorder (Stockman, 1996). Given that in the current study, a standardized protocol was used to elicit different types of semantic content categories during the freeplay session, it is suggested the absence of the identified semantic content categories tended to suggest PH's potential disruptions in producing them instead of a result of his lack of interests in expressing the corresponding semantic content categories in the freeplay session.

The different patterns in the expressions of semantic content categories observed in LY and PH possibly indicated that the manifestations in the disruptions in expressing semantic content categories are not always prominent among children with language difficulties. As such, the measures of utterance length, lexical diversity, and/or syntactic complexity seem to be better for the identification of individuals with language disorders (e.g., Hewitt et al. 2005; Klee, 1992; Rice et al., 2010; Watkins et al., 1995), compared with the profiling according to

semantic content categories produced. Nevertheless, the profiling according to semantic categories produced does serve a different purpose, which plays a more significant role in identifying subtle disruptions in language production that may not be easily captured using quantitative measures of utterance length, lexical diversity and/or syntactic complexity.

4.6.3 Manifestations of Disruptions in the Content-Form Interface

To elaborate on how language profiling plays a significant role in identifying subtle disruptions in language productions not easily captured in measures of utterance length and lexical diversity, more in-depth analyses at the content-form interface were conducted for further illustration. Specifically, the variety of content categories expressed using verbs and the diversity of lexical forms in expressing the content category of *temporal* by the two LD participants were analyzed.

Regarding the semantic content categories represented by verbs, it was expected that LY, at her age, would express three of the seven semantic content categories. However, the findings revealed that LY only expressed *action* and *locative action* categories using verbs, while the expression of the *state* category was absent. On the other hand, PH demonstrated the ability to use verbs to express all the age-expected semantic content categories. Notably, there was limited variation in the verb type used to express the *dative* category in his entire language sample. The above observations tended to suggest that the two studied children with language difficulties may demonstrate inadequacy in using a particular syntactic form to express sufficiently diverse semantic content categories, when compared with children of a similar age range.

On the other hand, the results revealed that both LY and PH had limited abilities to express *temporal* content with various lexical items, and relied solely on the use of aspect

115

markers. None of them were observed to use temporal adverbs or temporal nouns to convey temporal information. Additionally, the types of aspect markers used were relatively restricted. According to previous literature, the use of temporal adverbs is widely agreed to emerge between the ages of 24 and 30 months (e.g., Liang et al., 2019; Zhou, 2004). Besides, it is also reported in study two that temporal adverbs and temporal nouns, especially those more proximate and generic terms, could be produced by Mandarin-speaking children as young as 2 years old. The current observations therefore suggested that the two children with language difficulties my have demonstrated inadequacies in expressing a diversity of linguistic forms and lexical items to encode a particular semantic content category.

By investigating the content-form interface in children with language difficulties, that is, the expression of different semantic content categories with verbs, as well as the expression of semantic content category of *temporal* with different linguistic forms, the current study tended to show that children with language difficulties may encounter subtle disruptions in their expressions which may not be easily observed using only measures of syntactic complexity and language diversity. Children with language disorders often demonstrate language disruptions in the domains of content and form (Bishop, 1997; Leonard, 2014). Therefore, it is possible that the disruptions in the two domains interact and interfere with each other. The results of the current study appeared to echo a previous study that impaired semantic representations and syntactic difficulties interacted and the language disruptions manifested through the content-form interface in children with language disorders (Mok & Kipka, 2009). Consequently, the findings suggest that language profiling plays a significant role in identifying the diverse manifestations of disruptions observed in language disorders. This information is vital for effective intervention planning. By understanding the specific areas of disruption, such as limitations in semantic content categories and the content-form interface, clinicians can tailor intervention strategies and plan the specific goals to target the language needs of

individuals with language disorders. Language profiling thus provides valuable insights for developing individualized intervention plans to address the identified disruptions.

4.6.4 Possible Factors Affecting the Expression of Semantic Content Category in Children with Language Difficulties

In study one, it is proposed that cognitive complexity is one of the factors affecting the acquisition of various semantic content categories in typically developing Mandarin-speaking children. Cognitively more complex semantic content categories were found to be acquired later than those associated with lower cognitive complexity. It is proposed that children acquire semantic content categories along with the advancement in cognitive ability. Initially, children acquire concrete knowledge of objects and events through sensorimotor experiences. As their cognitive abilities advance, they gradually progress to understanding more specific and abstract relations between events. Finally, they develop the capacity for implicit and non-transparent reasoning processes, enabling them to grasp more complex semantic content categories.

Indeed, it is well-documented that children with language disorders often exhibit limitations in a variety of cognitive abilities. These include working memory abilities (Archibald & Gathercole, 2007; Montgomery & Evans, 2009) and processing speed (Leonard et al., 2007; Montgomery & Windsor, 2007). Deficits in executive functions, such as inhibiting interference from competing processes, have also been identified (Evans et al., 2018). Additionally, attentional control, which involves the ability to sustain attention and switch attention when necessary, has been found impaired in children with language disorders (Bishop & Norbury, 2005; Montgomery et al., 2009). In general, the cognitive impairments observed in children with language disorders can have a significant impact on their language processing abilities. These processing difficulties can interfere with their ability to efficiently access

117

language input and retrieve the acquired knowledge in a flexible manner, causing disruptions in various domains of language (Leonard et al., 2007).

Building upon this knowledge, it is reasonable to speculate that the disruptions observed in the expression of specific semantic content categories and limited vocabularies within those categories may be associated with their cognitive impairments. However, further studies are necessary to validate this hypothesis and explore the precise relationship between cognitive abilities and the difficulties encountered in expressing semantic content category among children with language disorders. Continued research in this area will contribute to a more comprehensive understanding of the underlying mechanisms and inform targeted interventions for children with language disorders.

4.7 Concluding Remarks

The current study examines the expressions of semantic content category and the content-form interface in two Mandarin-speaking children with language difficulties, by comparing the language samples between the two cases and children with typical development reported in study one. The results indicating shorter utterance length and lower lexical diversity in the two referred participants strongly suggest the presence of language disorder (e.g., Klee, 1992; Rice et al., 2010). Importantly, manifestations of disruptions in expressing semantic content category and the content-form interface were illustrated. It is proposed that documenting the semantic content category in the language productions of children with language difficulties serves multiple purposes. Firstly, it enriches the language profiles by providing a more comprehensive understanding of the specific semantic content categories that may be affected. This detailed analysis goes beyond traditional quantitative measures and allows for a deeper exploration of the child's language disruptions. Secondly, identifying subtle disruptions in semantic content categories

manifested through the content-form interface provides valuable insights that may not be easily captured through quantitative measures alone. Clinically, the observed disruptions in the semantic content category and the content-form interface offer valuable intervention directions, informing clinicians about additional areas of language disruptions that require support, so that the impact of language disorder can be mitigated.

Chapter Five

Summary and General Discussion

This thesis adopted the content-form framework (Bloom & Lahey, 1978) and analyzed the language profiles of Mandarin-speaking children, comparing those with and without language disorder in both content and form domains, as well as the interface between the two. The primary objective of the thesis is to imply intervention by investigating whether language disruptions can manifest in the expression of semantic content categories, and their corresponding content-form interface in Mandarin-speaking children with language disorder. To achieve this, the Corpus of Mandarin Child Language (CMCL) that documented the production of different semantic content categories by typically developing Mandarin-speaking children was established. An acquisition trajectory of the semantic content category was suggested. After that, the acquisition of the semantic content category of *temporal* was investigated thoroughly to illustrate the content-form interface in Mandarin-speaking children. In addition, language samples of two children with language difficulties were also compared with those of children with typical development to contribute additional understanding of the content domain and content-form interface of language disorder. By conducting this comparison, valuable insights are gained regarding the specific strengths and weaknesses exhibited by children with language disorders across various language domains, aiming to provide a foundation for developing comprehensive interventions that encompass specific treatment goals suitable for addressing individual needs.

The following reviews the major findings of the three studies and a general discussion on the overall findings can also be found after that.

5.1 Summary of Major Findings

Study one documented the language expressions of 82 native Mandarin-speaking children using a syntactically and semantically annotated database – CMCL. The results of the study aligned with previous research on typical language acquisition, specifically in terms of utterance length and lexical diversity. This finding confirmed the clinical value and usefulness of the data obtained from the CMCL. In addition, the study proposed an acquisition trajectory of semantic content categories in Mandarin-speaking children. Notably, the findings indicated that the acquisition pattern closely resembled that of English-speaking peers, as previously observed by Bloom (1991). The acquisition pattern observed in this study can be largely attributed to the cognitive and syntactic complexity associated with the semantic content categories. Additionally, the acquisition process in Mandarin is influenced by language-specific properties and cultural factors. Besides, the interaction between semantic content and language form in the acquisition process was also illustrated, by analyzing how a specific form (i.e., verb) was used to express different semantic content categories, and how different forms were employed to convey a particular semantic content category (i.e., *dative*). The findings revealed that verbs expressing contents of *action* and *state* emerged earlier compared to verbs expressing other contents like *notice* and *communication*. On the other hand, older children demonstrated a wider range of lexical items and syntactic structures when expressing the content of *dative* compared to younger children. These observations shed light on the relationship between semantic content and language form, emphasizing their impact on the language acquisition process in Mandarin-speaking children.

Further to the observations in the content-form interface in study one, study two examined a particular semantic content category – *temporal* – thoroughly for a more detailed investigation of the content-form interface in child language acquisition. First, to explore how

Mandarin-speaking children acquire different forms to express the same content, the acquisition between three different temporal-marker groups, namely aspect markers (AMs), temporal adverbs (TAs) and temporal nouns (TNs), by Mandarin-speaking children was examined. Results indicated that various time concepts proposed by Weist (1989), regulated the acquisition trend among the three different temporal marker groups. In particular, children mastered the use of AMs to express the internal time perspective of events before acquiring the *event time* concept. The subsequent acquisition of TAs to represent the past/non-past timeframe indicated the emergence of the *event time* concept, while the use of TNs to express a more external timeframe signaled the initial knowledge of the *reference time* concept. Finally, higher co-occurrences of temporal markers in single utterances found in older participants reflected the mastery of the *reference time* concept. As these time concepts affect the semantic representations of time, advancement in their knowledge provides the ingredients for expressing time and therefore affects the acquisition of different temporal-marker groups. Next, to examine how Mandarin-speaking children's acquisition of forms is affected by content, the acquisition within each group was also examined. Results indicated that the acquisition of different lexical forms within each temporal-marker group also appeared to be affected by the semantic features of temporal remoteness and specificity, with the more remote and specific ones being acquired later. It is suggested that knowledge of both content and form interact to modulate the acquisition of temporal markers in Mandarin.

Finally, considering the diverse manifestations of disruptions in children with language disorders, study three examined the language profiles of two Mandarin-speaking children with language difficulty in the domains of content and form. This investigation involved comparing their language production to the findings reported in study one, which focused on typically developing Mandarin-speaking children. Results on shorter MLU and lower lexical diversity in

the two participants verified their language difficulties and tended to suggest their likelihood of having language disorder (e.g., Klee, 1992; Rice et al., 2010). Besides, manifestations of disruptions in expressing different semantic content categories among children with language difficulty were found. Additionally, disruptions that lie in the content-form interface were also observed in their language productions. Notably, both participants tended to show limitations in the variety of semantic content categories expressed by a certain form, and the range of lexical items used to convey a specific semantic content category. By considering the concept of content-form interface, these findings provide implications for clinical intervention targeting children with language disorders.

5.2 General Discussion

5.2.1 Importance of Investigating Semantic Content Category and the Content-Form Interface in Child Language Disorder

Investigating semantic content categories and the content-form interface in child language disorders holds significant importance for several reasons. Firstly, an acquisition trajectory of various semantic content categories by Mandarin-speaking young children was proposed. Remarkably, the acquisition trends of semantic content category observed in typically developing Mandarin-speaking children align with previous studies on the acquisition of semantic content categories by English-speaking children, such as Bloom (1991). This not only provides additional cross-linguistic evidence but also highlights the language-general nature of these acquisition patterns, which constitutes an important basis of studying language disorders.

Besides, examining the expressions of semantic content categories in children with language disorders enriches our understanding of their language profiles from a semantic

123

perspective. In particular, the manifestations of disruptions in expressing certain semantic content categories and the content-form interface observed in the thesis tended to suggest the potential difficulties in acquiring semantic content categories in children with language disorders. This finding was also in line with a previous study which reported fewer types of semantic content categories produced by children with language disorder (Stockman, 1996), and again supported that the content domain can also be impaired in language disorder. By going beyond traditional measures such as mean length of utterance and lexical diversity, an additional focus on the semantic content category provides supplementary information in child language. This approach allows for the identification of relative strengths and weaknesses across different language domains, contributing to a comprehensive assessment of language abilities in children with language disorders.

Thirdly, certain semantic content categories may be more susceptible to disruptions in children with language disorders. It was found that semantic content categories related to complex sentences, or associated with higher cognitive complexity would fall behind those related to simple sentences, or those cognitively less complex ones respectively in the acquisition process. Semantic content categories that are associated with higher cognitive and syntactic complexity tend to place greater demands on cognitive abilities such as working memory and processing speed. Understanding the challenges faced in expressing these semantic content categories informs intervention strategies and helps address the specific linguistic difficulties experienced by children with language disorders.

Lastly, disruptions in semantic content category manifested through the content-form interface were indicated in study three. One of the participants with language difficulty was found to be able to express all the age-expected semantic content categories with no apparent disruptions. However, detailed investigations revealed that the semantic content categories

expressed by certain forms, as well as the diversity of lexical items expressing certain semantic content categories were both limited. These potential impairments may not be easily discovered if only a single language domain is examined. These findings appear to suggest that disruptions in semantic content categories may not always be prominent and easily captured by traditional measures. These subtle disruptions, as a result, may go unnoticed using conventional assessment methods. A focus on the semantic content category and its content-form interface may therefore reveal nuanced language difficulties in children with language disorders.

In summary, investigating semantic content category and the content-form interface in child language disorders has multiple benefits. It adds to our understanding of language acquisition patterns, enriches language profiles by considering the content domain, identifies semantic content categories more prone to disruptions, and uncovers subtle disruptions that may require tailored interventions. This line of research contributes to a holistic understanding of language disorders and facilitates the development of effective therapeutic interventions for affected children.

5.2.2 Clinical Implications for the Intervention of Children with Language Disorder

Current findings echo previous literature that disruptions can manifest in the expressions of semantic content category (Stockman, 1996), or the interface between content and form in language disorder (Mok & Kipka, 2009). As an integral part of the intervention, it is crucial to conduct a comprehensive assessment to identify the areas of disruptions and plan for intervention. To this end, traditional clinical assessment for children with language disorders has primarily focused on examining the language form (Bååth1 et al, 2019) and assessment in the area of semantics is often limited in measuring vocabulary size only, which may not be adequate in showing a complete picture on the deficit in the semantic domain (Brackenbury &

Pye, 2005). In addition, since the manifestations of language disruptions in different domains vary, documenting a single language domain seems to be insufficient to justify the profiling. As such, language profiling based on language samples provides a useful procedure for describing language at different linguistic domains and levels, so that individuals exhibiting different types of language difficulties can be documented (Long, 2012). This also provides a platform for us to investigate how children with language disorders integrate their knowledge of different language domains. For instance, it may be useful to understand how the production of a particular content category is affected by the deficit in constructing the associated syntactic forms, or conversely, how the production of one specific form can be affected by the deficit in establishing the semantic representation of a particular semantic content category. Moreover, this clinical procedure allows a thorough evaluation of an individual's relative strengths and weaknesses, and provides a basis for a more thorough and in-depth remedial intervention (Crystal, 1982).

Given the importance to examine the content domain in young children with language disorder, it would be beneficial to include an assessment of semantic content category. As discussed before, language sample analysis (LSA) makes use of naturalistic child language data is regarded as an ecologically valid and authentic method of assessment and can be used to assess children's language across different domains (Owens, 2010). Conventionally, measures of utterance length and lexical diversity are used in the language profiling of children with language disorders. While content domain should also be considered in the assessment of language disorder, the language sample analyses need to move beyond those measures of surface structures and utilize linguistic frameworks capable of addressing the interaction between content and form (Mok & Kipka, 2009). Examining the semantic content category thus has the potential to complement traditional analyses on form or word-level semantics. To

achieve this, it is recommended to design specific procedures and special context for eliciting various semantic content categories (e.g., presentation of undesired items to elicit *reject*; use of story materials to elicit *temporal* or *causal* in narratives) during the language sample taking procedures. After that, the language samples can be annotated with tags on semantic content category in addition to traditional annotations on part-of-speech (MacWhinney, 2000; Sagae et al. 2010). Although it can be quite time-consuming for clinicians to get familiar with and to conduct analyses on the semantic content category, the information from the semantic perspective will be of great importance for understanding and enriching the language profiles of children with language disorders.

Apart from comprehensive assessment and profiling of children's language abilities, intervention for children with language disorders also encompasses treatment provision on the identified area of disruptions. The treatment aims to enhance children's abilities in specific aspects of language and utilize their linguistic strengths to facilitate overall learning (Van der Lely, 1993). As noted before, there is a possibility that children with language disorders exhibit disruptions in expressing various semantic content categories. Therefore, intervention in this area should also be explored, along with the existing intervention options. Moreover, the potential manifestations of disruption in content-form interface may also call for specific intervention considerations. Indeed, previous literature has suggested intervention working on both content and form together (Ebbels et al. 2007; Mok & Kipka, 2009). Mok and Kipka (2009) recommended that intervention should involve the establishment of both correct semantic representations and appropriate syntactic structures together. Likewise, interventions using an integrated approach have also been reported for children with language disorders (Ebbels et al., 2007). This study investigated the efficacy of different treatment options for students with language disorders using randomized control trials. Twenty-seven students with language

disorders were randomly assigned to the syntactic-semantic therapy, semantic therapy or control group. Results indicated that both therapies were effective in improving the participants' ability to produce obligatory verb argument structures, and the syntactic-semantic therapy additionally increased the use of optional verb arguments. The progress from both therapies was also found to be generalized and maintained. On the other hand, no significant progress was made in the control group. This study not only evidenced the interaction between impaired semantic representations and syntactic difficulties in language disorder, but also pioneered an intervention approach working on domains of both form and content simultaneously.

In light of the above, when designing language targets for intervention, both form and content should be considered, with the disruptions in both domains being documented in an enriched language profile. A two-dimensional intervention approach similar to the proposal on 'new forms express old functions and new functions are expressed by old forms' (Slobin, 1973) is hereby suggested. While the treatment objective is to enhance children's overall language abilities in both domains of content and form, new information will only be introduced to one domain each time in the intervention. For instance, treatment can be designed to establish various semantic content categories expressed with an existing form. Alternatively, a variety of forms being used to express an existing semantic content category can also be expanded. It may also be useful to find out particular syntactic or semantic cues for the intervention so that the children's relative strength can be made use of to improve the area of weaknesses in the remediation.

The findings in study three indicated a high risk of language disorder in the two participants under investigation. Intervention is needed to address their language difficulties that appear in various domains. To address the apparent disruptions in expressing semantic content categories, it is essential to conduct a more comprehensive assessment to identify the

specific difficulties. Besides, it is also recommended to adopt a child-centered treatment approach that fosters a natural and facilitative environment. This approach aims to create an optimal learning context that supports the child's acquisition of various semantic content categories and addresses their specific needs. In the following, the intervention plans for LY and PH are presented to illustrate how the abovementioned two-dimensional framework is adopted in setting up the intervention program.

5.2.2.1 Intervention Planning for LY

LY, a 3;11 girl reported to have language difficulty, demonstrated shorter MLU and lower NDW when compared with the TD age-matched peers reported in study one. In particular, her production of all open class words (i.e., noun, verb, adjective) and closed class words was observed to exhibit lower diversity. Considering the content domain, LY was able to exhibit the three age-expected semantic content categories, and even produced the content of *dative* which was expected in the 4 year olds. No disruptions in expressing semantic content categories were suggested from the above preliminary results. However, detailed investigations into the content-form interface indicated that she did not express the content of *state* with verbs, and the types of verbs expressing the content of *locative action* were also limited in her language sample. Moreover, the types of temporal markers used to express *temporal* content were also restricted to aspect markers only. To address all the above inadequacies, it is recommended that the intervention direction should focus on expanding the length and variety of syntactic structures and vocabulary expansion (e.g., Balthazar & Scott, 2017; Lau et al., 2023; Pomper et al., 2022), with the consideration of the content-form interface.

i.   Old Content New Form

In expanding the length and variety of various syntactic structures, a content-driven intervention for establishing new language form is suggested. Novel syntactic structures could be introduced to express a range of semantic content categories that have already emerged in LY (e.g., *locative state, possessive*). This makes use of her strength in the content domain in facilitating her weakness in syntactic form.

Next, it is recommended to expand the limited types of temporal markers used by LY to express the content of *temporal*. According to the literature, four different aspect markers are commonly used by Mandarin-speaking children and they are usually acquired before the age of two (e.g., Zhou, 2004). Besides, temporal adverbs and temporal nouns are other linguistic devices frequently employed to express temporality in Mandarin (Tse et al., 2012). To expand the variety of lexical forms in expressing this established content, more different aspect markers (e.g., experiential 过 *guo4*), or more different temporal markers groups such as temporal adverbs and temporal nouns, can therefore be introduced in the intervention of the child.

ii.   Old Form New Content

Concerning the impairment in lexical diversity, fewer types of verbs were observed in LY's production as compared with age-matched peers. Here a goal-selection approach based on semantic content category is proposed, in which new semantic content categories expressed by verbs should be introduced. For LY, the semantic content categories of *state* and *locative action* can be introduced to expand her verb inventory, as the two were either absent or expressed with restricted lexical items and appeared to be her language weakness.

5.2.2.2 Intervention Planning for PH

PH, a 4;02 boy referred for having language difficulty, also demonstrated shorter MLU and lower NDW when compared with the TD age-matched peers reported in study one. Specifically, his production of verbs, adjectives and closed class words was observed to exhibit lower diversity. Considering the content domain, three age-expected semantic content categories were missing in his language samples, namely *denial*, *additive* and *causal*. The above results indicated his weaknesses in the domains of both content and form. To address these disruptions, it is recommended that the intervention direction should focus on expanding the length and variety of various syntactic structures, vocabulary expansion, as well as introducing the missing semantic content categories. The content-form interface should also be considered when working out the detailed treatment objectives.

i.   Old Content New Form

First, similar to the intervention of LY, a content-driven approach for establishing new language form is suggested to expand the utterance length and variety of different syntactic structures in PH, by considering the semantic content categories which have already emerged. For example, a more detailed look at the production of *dative* content revealed that only one single utterance using the verb 给 'give' in a simple subject-predicate structure (姐姐给弟弟汉堡包 'sister gives a burger to the younger brother') was found in the child's language sample. In order to expand the lexical items expressing this *dative* content, other different verbs (e.g., 帮，送) or prepositions (e.g., 让，对) reported in study one may also be introduced. Besides, other syntactic structures like serial verb construction (e.g., 帮他捡起来 'help him pick up') and pivotal sentences (e.g., 请你吃汉堡包 'treat you a burger') were also reported in study

one, which may also be targeted so that the variety of sentence structures expressing the content of *dative* can be increased in PH.

Next, the initially established semantic content category, *temporal*, was also found to be expressed with the perfective aspect marker 了 *le* only. Same as the intervention for LY, other types of aspect markers, or more different temporal markers groups (i.e., temporal adverbs and temporal nouns), can also be introduced in the intervention of PH in expressing the *temporal* content.

ii.   *Old Form New Content*

Three semantic content categories, *denial, additive* and *temporal*, were expected but found missing in PH's production. Considering the content-form interface, the absent age-expected semantic content categories can be established by utilizing existing forms in the child's production. For instance, the content of *denial* can be introduced with an existing state verb (e.g., 喜歡 'like') by adding a negative marker 不 'no'. Besides, the content of *additive* can be established by encouraging the child to connect two simple sentences that he has no difficulties producing (e.g., 哥哥刷牙 'Brother brushes teeth.', 哥哥洗脸 'Brother washes face'. Finally, the content of *causal* may be established at a later stage. Once he can connect two sentences and achieve a variety of temporal expressions, the content of *causal* can be introduced using existing lexical items and syntactic structures (e.g., 爸爸口渴了 'Dad is thirsty.', 他喝果汁 'He drinks juice.')

The above recommendations demonstrate how the content-form interface can be considered in designing the specific goals for children with language disorders. By analyzing the domains of both content and form in the language samples, the areas of strength and

weakness of the targeted child can be identified. Intervention direction is guided by a balanced consideration of the disruptions in both content and form, which not only allowed more directions in examining the language disruptions, but enriched the language profile in documenting the heterogenous language manifestation in children with language disorders. In addition, this framework also contributes toward determining the best possible interventions for children with language disorders, based on a detailed and comprehensive assessment of the domains of content and form, as well as the content-form interface.

5.3 Significance of Findings

The current studies reported in this thesis examined the language profiles of Mandarin-speaking children with or without language disorders, in the domains of content and form, as well as the content-form interface. The findings have contributed significance empirically, theoretically and also clinically.

5.3.1 Empirical Novelties

Study one presented empirical data on the acquisition of various semantic content categories by typically developing Mandarin-speaking children. This is the first study that investigated the notion of semantic content category in Mandarin. The results contributed experimental data on the acquisition of semantic content categories in Mandarin-speaking children and suggested that the acquisition trajectory resembled that of English-speaking children. In addition, as mentioned before, acquisition studies in child language have focused primarily on language form or lexical semantics, leaving the content domain beyond lexical

level relatively unattended. Findings from this study therefore contributed to additional understanding of child language acquisition from the semantic perspective.

Given the importance of content domain in child language acquisition (e.g., Mok & Kipka, 2009), the current study adopted language sample analysis in examining various semantic content categories produced by the children. A specially designed protocol was used to elicit the production of various semantic content categories in children's naturalistic speech. As such, samples obtained from this authentic method of assessment (Owens, 2010) can be utilized to document the content domain in addition to syntactic form. The CMCL, which possessed annotations of both form (part-of-speech) and content (semantic content category), was subsequently established. As mentioned before, currently available Mandarin corpora were mainly annotated syntactically (e.g., Deng & Yip, 2018; Li & Zhou, 2008). The CMCL thus provides an important experimental platform for future research examining the content domain, as well as the content-form interface in child language study. This also serves as an important basis to understand language disorders.

Regarding the particular semantic content category of *temporal*, previous research has predominantly concentrated on individual temporal-marker groups, such as aspect markers (e.g., Li & Bowerman, 1998) or temporal adverbs (e.g., Liang et al., 2019). In study two, a different approach was taken by exploring the acquisition of all three temporal marker groups simultaneously. This unique investigation provided empirical data on the overall acquisition of diverse forms used to express temporality. In addition, by investigating the semantic features of temporal remoteness and specificity within each of the temporal marker groups, the content-form interface was also explored, which further enriched the literature on the acquisition of *temporal* content.

Similar to the studies on child language acquisition, literature on child language disorder has also focused extensively on language form. Only limited studies have documented the language disruptions in the content domain, especially for the difficulties beyond lexical level (e.g., Alt & Plante, 2006). As such, study three reported two cases of Mandarin-speaking children with language difficulties, and provided empirical evidence on the potential disruption of semantic content category in children with language disorder. In addition, it was also indicated that the disruptions in content can also be manifested through the content-form interface. The above contributes novel empirical findings in child language disorder.

5.3.2 Theoretical Significance

The current studies have enriched the literature on the acquisition of semantic content categories in young children, as well as the interaction between content and form. Theoretically, a two-dimensional framework for language acquisition was suggested. When children acquire language in the early years, it is proposed that the language-general semantic content and the language-specific syntactic form interact in a bidirectional manner, and this content-form interface plays an essential role in regulating the language acquisition of young children. The findings contributed additional understanding of language acquisition in the semantic domain, as well as the content-form interface. Additionally, the study of language disorders is also importantly implied.

In particular, study two proposed that the advancement in children's time concepts (Weist, 1989) affected the acquisition between different temporal-marker groups, while the semantic features of temporal remoteness and specificity also altered the acquisition of lexical items within each temporal-marker group. In line with Weist's (1989) hypothesis on the universal constraint of time concepts on the acquisition of linguistic forms, this study adds

crosslinguistic evidence that the time concepts provide language-general building blocks for children to use linguistic-specific forms to code time. It is further suggested that this content-form interactive framework can also be applied to explain the acquisition of other semantic content categories.

5.3.3 Clinical Significance

In light of the findings in study three, where potential disruptions in expressing certain semantic content categories were identified, it is useful to consider assessing the expressions of these semantic content categories in children with language disorders. By examining and documenting the expressions of various semantic content categories in children with language disorder, the language profiles can be enriched from the semantic perspective. Eventually, a deeper understanding of the heterogeneous language manifestations observed in language disorders can also be obtained. Furthermore, the two-dimensional content-form interactive framework proposed earlier has significant implications for the assessment and treatment of children with language disorders. It emphasizes the importance of thoroughly evaluating both the content and form domains in order to identify areas of disruptions and understand the interaction between the two. This comprehensive evaluation also allows for the identification of relative strengths and weaknesses in language abilities, which is crucial for effective treatment planning.

On the other hand, balanced considerations in both content and form are also recommended in the treatment planning for children with language disorders. By addressing both content and form, intervention programs can ensure a well-rounded and comprehensive approach to language acquisition, enabling children to communicate effectively in various contexts. Children with language disorders may also experience disruptions manifested in the

content-form interface, leading to difficulties in accurately conveying their intended meaning. By addressing both content and form, intervention programs can strengthen the connection between meaning and language structures, enhancing the child's ability to express themselves accurately and effectively. Lastly, children with language disorders often exhibit variations in their strengths and weaknesses in different language domains. By considering both content and form, intervention programs can identify the specific areas of difficulty for each child and tailor the intervention accordingly. This personalized approach ensures that the child's unique needs are addressed, maximizing the effectiveness of the intervention.

5.4 Limitations of the Study

This thesis examines the acquisition of semantic content categories, as well as the content-form interface, in Mandarin-speaking children with or without language disorder. Several limitations of the studies were identified. First, when examining the acquisition of various semantic content categories, some of them such as *specification* and *epistemic* may be too complex and abstract to acquire at an early age (Bloom, 1991; Evers-Vermeul & Sanders, 2011). Similarly, in study two, the repertoire of distant TAs observed in children aged two to four was smaller than that of proximate TAs, but this might only imply that the limited range of distant TAs used by children of these ages impeded us from testing the effect of temporal remoteness on the acquisition of temporal markers. Given that – for both linguistic and cognitive reasons – it takes more time for a preschooler to fully master distal and abstract concepts, quite a few TAs were simply not used by the preschoolers we worked with. In view of the above, future research should therefore be extended to include older children, so as to

gain a clear overall picture of the usage and acquisition of various semantic content categories, as well as how various temporal markers are acquired and used.

On the other hand, LSA adopted in the current studies provides a rich communicative context and allows naturalistic language production in children (Evans & Craig, 1992; Owens, 2010). However, as proposed in study two, some semantic content categories such as *recurrence* and *reject* may be culturally sensitive and cannot be elicited within the context provided. It may be necessary in future studies to include some structural language-eliciting procedures, to capture some particular semantic content categories.

Other limitations of the present work with typically developing children include its cross-sectional nature and the uneven number of subjects in each age group. A sample with more even distribution of subjects across ages, and a longitudinal study with longer observation windows would be preferable in the future.

Regarding the investigations in children with language disorders, the case studies in study three shed new light on the clinical application of semantic content categories and the content-form interface in the intervention. However, the small number of participants with language difficulty may not allow in-depth analysis of the underlying disruption. As heterogeneity exists among children with language disorders (Bishop, 1997; Leonard, 1998), it may be necessary for future studies to include a larger sample size of children with language disorders. Specific experimental designs should also be adopted to find out the significant differences between the two groups of children in the acquisition of various semantic content categories. Besides, the two participants were referred by a local therapist as they were reported to encounter language difficulties that affect their daily functioning. Although shorter utterance length and lower lexical diversity were observed and tended to suggest the existence of language disorder, it will be better to confirm their language difficulties and determine the

severity. Replication of the current study, with the conduction of standardized assessment to verify participants' language status, is therefore suggested.

5.5 Implications for Future Work

In this thesis, the CMCL, a child language database annotated both with parts-of-speech and semantic content categories was established, which provides a useful and convenient tool for us to study the language acquisition of young children from both syntactic and semantic perspectives. While the acquisition of various temporal markers in expressing the semantic content category of *temporal* was investigated in study two, similar in-depth investigations on the acquisition of other semantic content categories in Mandarin are also inspired and supported. Importantly, the two-dimensional framework with content-form interaction on language acquisition is proposed in this thesis. It remains unclear whether such framework can also predict the acquisition of other semantic content categories. That is, when acquiring a particular form, children progressively broaden the semantic content categories being expressed by that form, and similarly, children also gradually expand the use of different lexical items and syntactic structures to represent that semantic content category when acquiring a particular semantic content category. Future studies can be undertaken to validate the aforementioned findings.

In future studies, detailed investigation on the effect of cognitive and syntactic complexity between and within each of the semantic content category on their acquisition can be conducted to advance our understanding on the acquisition trend. Besides, it is important to take into account not only a larger sample size and inclusion of older participants, but also the socioeconomic status (SES) of the participants on the acquisition of semantic content

categories. Considering SES is crucial because it has been demonstrated to influence the assessment of various areas within the content domain (e.g., Brakenbury & Pye, 2005; Dollaghan et al., 1999). Therefore, incorporating SES as a variable in future studies would provide a more comprehensive understanding of how socioeconomic factors influence the acquisition of semantic content categories. On the other side, language input and feedback from caregiver and peers also play a crucial role in language acquisition (Pinker, 1984). These environmental inputs may help to refine the understanding of various semantic content categories and therefore maybe investigated further for their influence on acquisition of semantic content category.

Besides, future research could also aim at investigating the vulnerability of various semantic content categories in language disorder. Additionally, investigating the effectiveness of various intervention approaches that focus on a single domain, or domains of both content and form can also be conducted. Finally, while the domains of content and form, as well as their interface were the main focus of this thesis, the third language domain of language use may also be investigated to examine the complex interplay between these three language dimensions in language disorder.

5.6 Conclusion

Language disorders have a profound impact on children's communication skills and daily functioning, emphasizing the crucial need for effective intervention plans to mitigate negative consequences. This thesis aimed to investigate the language profiles of Mandarin-speaking children, both with and without language disorders, focusing on the domains of content and form. Additionally, it explored the content-form interface and examined potential

disruptions in expressing semantic content categories among children with language difficulties. The findings of this thesis significantly contribute to our understanding of the language profiles of Mandarin-speaking children with or without language disorders, particularly from a semantic perspective and concerning the content-form interface. They provide valuable experimental data on the acquisition of semantic content categories in Mandarin-speaking children. The establishment of the Corpus of Mandarin Child Language (CMCL) offers a platform for examining the content domains and the content-form interface in early child language acquisition. Theoretical implications suggest a bidirectional interaction between language-general content and the language-specific form, regulating the language acquisition of young children. The findings regarding children with language disorders shed light on the potential manifestations of disruptions in expressing semantic content categories and the content-form interface in language disorders, offering new perspectives on profiling language abilities across various language domains for children with language disorders. Moreover, the study highlights the importance of considering a balanced approach to intervention and goal planning for children with language disorders, taking into account both content and form, as well as their interface, ultimately improving their communication skills and quality of life.

References

Addis, D., & Schacter, D. (2008). Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus, 18*(2), 227-237

Adolph, K. E., Berger, S. E., Kuhn, D., & Siegler, R. S. (2006). Handbook of child psychology: Vol. 2. C*ognition, perception, and language*.

Aguado-Orea, J., & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *PloS one, 10*(3), e0119613.

Alink, L. R., Mesman, J., Van Zeijl, J., Stolk, M. N., Juffer, F., Koot, H. M., ... & Van IJzendoorn, M. H. (2006). The early childhood aggression curve: Development of physical aggression in 10-to 50-month-old children. *Child development*, *77*(4), 954-966.

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*(5), 941–954. https://doi.org/10.1044/1092-4388(2006/068)

Ambridge, B., & Rowland, C. F. (2013). Experimental methods in studying child language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(2), 149-168.

Archibald, L. M., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic bulletin & review, 14*(5), 919-924.

Arndt, K. B., & Schuele, C. M. (2013). Multiclausal utterances aren't just for big kids: A framework for analysis of complex syntax production in spoken language of preschool- and early school-age children. *Topics in Language Disorders*, *33*(2), 125-139.

Bååth, R., Sikström, S., Kalnak, N., Hansson, K., & Sahlén, B. (2019). Latent semantic analysis discriminates children with developmental language disorder (DLD) from

children with typical language development. *Journal of psycholinguistic research*, *48*, 683-697.

Balthazar, C. H., & Scott, C. M. (2017). Complex sentence intervention. *Treatment of language disorders in children*, *2*, 349-388.

Barrett, M. (1999). An introduction to the nature of language and to the central themes and issues in the study of language development. *The development of language*, 1-24.

Bi, H. Y., & Peng, D. L. (2002). Experiment on the comprehensive ability and strategies about three kinds of time adverb of 4-6-year-old children. *Psychological Development and Education, 4*, 22-25 [In Chinese: 毕鸿燕, & 彭聃龄. (2002). 4~6 岁儿童三种时间副词理解能力及策略的实验研究[J]. *心理发展与教育, 4,* 22-25].

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge University Press.

Biq, Y., & Huang, C. (2016). Adverbs. In C. Huang & D. Shi (Eds.), *A reference grammar of Chinese*. Cambridge University Press.

Bishop, D. V. (1997). Cognitive neuropsychology and developmental disorders: Uncomfortable bedfellows. T*he Quarterly Journal of Experimental Psychology Section A, 50*(4), 899-923.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., Baird, G., Bauer, A., Bellair, J., Boyle, C., Brownlie, E., Carter, G., Clark, B., Clegg, J., Cohen, N., Conti-Ramsden, G., Dockrell, J., Dunn, J., Ebbels, S., … Grist, M. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. Journal of Child Psychology and Psychiatry, 58(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bishop, D. V., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry*, 31(7), 1027-1050.

Bishop, D. V., & Norbury, C. F. (2005). Executive functions in children with communication impairments, in relation to autistic symptomatology: I: Generativity. *Autism, 9*(1), 7-27.

Bishop, D. V., Chan, J., Adams, C., Hartley, J., & Weir, F. (2000). Conversational responsiveness in specific language impairment: Evidence of disproportionate pragmatic difficulties in a subset of children. *Development and Psychopathology, 12*(2), 177-199.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, *35*(3), 619-650.

Bloom, L. (1970). *Language development; form and function in emerging grammars.* M.I.T. Press.

Bloom, L. (1991). *Language development from two to three.* Cambridge University Press.

Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. New York: Wiley.

Bowerman, M. (1973). Structural relationships in children's utterances: syntactic or semantic? *Cognitive development and acquisition of language*, 197-213.

Brackenbury, T., & Pye, C. (2005). Semantic deficits in children with language impairments: issues for clinical assessment. *Language, Speech & Hearing Services in Schools, 36*(1), 5–16. https://doi.org/10.1044/0161-1461(2005/002)

Branigan, H. P., Pickering, M. J., McLean, J. F., & Stewart, A. J. (2006). The role of local and global syntactic structure in language production: Evidence from syntactic priming. *Language and Cognitive Processes*, *21*(7-8), 974-1010. https://doi.org/10.1080/016909600824609

Brown, R. (1973). *A first language: the early stages.* Cambridge, Mass.: Harvard U.P.

Brownlie, E. B., Beitchman, J. H., Escobar, M., Young, A., Atkinson,L., Johnson, C., Wilson, B., & Douglas, L.(2004). Early lan-guage impairment and young adult delinquent and aggressive behavior. *Journal of Abnormal Child Psychology, 32*(4), 453–467.https://doi.org/10.1023/B:JACP.0000030297.91759.74

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech, Language, and Hearing Research, 36*(5), 948-958.

Chang, H. W. (1992). The acquisition of Chinese syntax. In *Advances in psychology* (Vol. 90, pp. 277-311). North-Holland.

Cheung, H. (2009). Grammatical characteristics of Mandarin-speaking children with specific language impairment. *Language disorders in speakers of Chinese*, 33-52.

Cheung, H. (1998). *Utterance length and the development of Mandarin Chinese*. Paper presented at The First Asia Pacific Conference on Speech Language and Hearing, 1998, Hong Kong.

Chung, T. T. R., & Gordon, P. (1998). The acquisition of Chinese dative constructions. In *Proceedings of the Boston University Conference on Language Development* (Vol. 22, pp. 109-120).

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65-110). New York: Academic Press.

Conti-Ramsden, G., & Botting, N.(2008). Emotional health in adolescents with and without a history of specific language impairment (SLI).The *Journal of Child Psychology and Psychiatry, 49*(5), 516–525.

Cox, M. K., & Key, C. H. (1993). Post hoc pair-wise comparisons for the chi-square test of homogeneity of proportions. *Educational and Psychological Measurement*, *53*(4), 951-962.

Crowe, K., & McLeod, S. (2020). Children's English consonant acquisition in the United States: A review. *American Journal of Speech-Language Pathology*, *29*(4), 2155-2169.

Crystal, D. (1982). *Profiling Linguistic Disability*. London: Edward Arnold.

Crystal, D., Fletcher, P., & Garman, M. (1989). *Grammatical analysis of language disability* (2nd Ed.). London: Cole and Whurr.

Cummings, L. (2009). *Clinical pragmatics*. Cambridge University Press.

de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, *2*(3), 267-278.

Demuth, K. (2015). Intonational development in early language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(3), 257-271.

Demuth, K., & McCullough, E. (2009). The acquisition of auxiliary syntax: A longitudinal elicitation study. *Journal of Child Language, 36*(1), 83-108.

Deng, X., & Yip, V. (2018). An event-semantic account of spatial modification in the acquisition of Mandarin. *Journal of Child Language*, *45*(4), 829–857.

Deng, X., Mai, Z., & Yip, V. (2018). An aspectual account of ba and bei constructions in child Mandarin. *First Language, 38*(3), 243-262.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: Cambridge University Press.

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N. & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, 42, 1432–1443.

Durrleman, S., Chen, L., & He, X. (2024). Acquisition of Mandarin long passives by children with developmental language disorder. *Clinical Linguistics & Phonetics, 38*(3), 260-284.

Ebbels, S. H., van der Lely, H. K. J., & Dockrell, J. E. (2007). Intervention for Verb Argument Structure in Children With Persistent SLI: A Randomized Control Trial. *Journal of Speech, Language, and Hearing Research, 50*(5), 1330–1349.

Ehrler, D. J., & McGhee, R. L. (2008). *PTONI: Primary test of nonverbal intelligence.* Austin, TX: Pro-ed.

Erbaugh, M. (1992). The acquisition of Mandarin. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 3, pp. 373-455). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ernst, T. (2001). *The syntax of adjuncts*. Cambridge University Press.

Esteve-Gibert, N., & Prieto, P. (2014). The development of prosodic boundaries in Spanish infants. *Journal of Child Language, 41*(4), 907-934.

Evans, J. L., & Craig, H. K. (1992). Language sample collection and analysis: interview compared to freeplay assessment contexts. *Journal of Speech, Language, and Hearing Research, 35*, 343-353.

Evans, J. L., Gillam, R. B., & Montgomery, J. W. (2018). Cognitive predictors of spoken word recognition in children with and without developmental language disorders. *Journal of Speech, Language, and Hearing Research, 61*(6), 1409-1425.

Evers-Vermeul, J., & Sanders, T. (2011). Discovering domains - on the acquisition of causal connectives. *Journal of Pragmatics*, *43*(6), 1645-1662.

Fang, H., & Hengeveld, K. (2022). Sentence-Final Particles in Mandarin. *Studia Linguistica*, *76*(3), 873-913.

Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego, CA: Singular Publishing Group.

Fey, M. E. (1986). *Language intervention with young children.*

Fu, M. Y. (2002). Adverbs of child language. Master's thesis, Anhui Normal University, Anhui, China [In Chinese: 傅满义. (2002). *儿童语言中的副词*(硕士学位论文). 安徽师范大学].

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200-213.

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*(4), 439-454.

Gentner, D. (1982). *Why nouns are learned before verbs: Linguistic relativity versus natural partitioning*. Center for the Study of Reading Technical Report; no. 257.

Gillam, R. B., & Marquardt, T. P. (2016). *Communication sciences and disorders: from science to clinical practice*. Jones & Bartlett Learning.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*(1), 3-55.

Goodglass, H., & Kaplan, E. (1972). *Boston Diagnostic Aphasia Examination (BDAE)*. Philadelphia: Lea & Febiger.

Grant, J. B., & Suddendorf, T. (2011). Production of temporal terms by 3-, 4-, and 5-year-old children. *Early Childhood Research Quarterly*, *26*(1), 87-95.

Green, M. (1975). Semantic acquisition in language impaired and normal speaking four year old children. *South African Journal of Communication Disorders*, *22*(1), 3–21.

Grela, B., Rashiti, L., & Soares, M. (2004). Dative prepositions in children with specific language impairment. *Applied Psycholinguistics*, *25*(4), 467-480.

Haebig, E., Kaushanskaya, M., & Ellis Weismer, S. (2015). Lexical processing in school-age children with autism spectrum disorder and children with specific language impairment: The role of semantics. *Journal of Autism and Developmental Disorders, 45*, 4109-4123.

Hao, Y., Sheng, L., Zhang, Y., Jiang, F., De Villiers, J., Lee, W., & Liu, X. L. (2018). A narrative evaluation of Mandarin-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 61*(2), 345-359.

Hao, M., Shu, H., Xing, A., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods, 40*(3), 728-733.

Herrmann, D. J., & Raybeck, D. (1981). Similarities and differences in meaning in six cultures. *Journal of Cross-cultural Psychology*, *12*(2), 194-206.

Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of Communication Disorders, 38*(3), 197-213.

Horvath, S., Kueser, J. B., Kelly, J., & Borovsky, A. (2022). Difference or delay? Syntax, semantics, and verb vocabulary development in typically developing and late-talking toddlers. *Language Learning and Development*, *18*(3), 352-376.

Horvath, S., Rescorla, L., & Arunachalam, S. (2019). The syntactic and semantic features of two-year-olds' verb vocabularies: A comparison of typically developing children and late talkers. *Journal of Child Language, 46*(3), 409-432.

Hsu, H.-Y., & Bishop, D. V. M. (2020). Semantic deficits in children with developmental language disorder: What do they reveal about the role of semantics in grammar? *Journal of Speech, Language, and Hearing Research, 63*(10), 3293-3306. https://doi.org/10.1044/2020_JSLHR-19-00269

Huang, C. C. (2003). Mandarin temporality inference in child, maternal and adult speech. *First Language*, *23*(2), 147-169.

Huang, R. L., Fletcher, P., Zhang, Z., Liang, W., Marchman, V., & Tardif, T. (2022). Early grammatical marking development in Mandarin-speaking toddlers. *Developmental Psychology*, *58*(4), 631.

Jarrold, C., Thorn, A. S., & Stephens, E. (2009). The relationships among verbal short-term memory, phonological awareness, and new word learning: Evidence from typical development and Down syndrome. *Journal of experimental child psychology*, *102*(2), 196-218.

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*(5), 1280–1299.

Jiménez, E., Haebig, E., & Hills, T. T. (2021). Identifying areas of overlap and distinction in early lexical profiles of children with autism spectrum disorder, late talkers, and typical talkers. *Journal of Autism and Developmental Disorders, 51*, 3109-3125.

Jin, Z. J. & Jin, X. M. (2008). Research on mean length of utterance and lexical diversity in preschool Mandarin-speaking children. *Chinese Journal of Evidence Based Pediatrics, 3*(4), 261-266. [In Chinese: 金志娟, & 金星明. (2008). 学龄前儿童普通话平均句子长度和词汇广度研究. *中国循证儿科杂志*, *3*(4), 261-266].

Johnston, J. R. (1985). Cognitive prerequisites: the evidence from children learning English. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, Vol. 1. The data; Vol. 2. Theoretical issues* (pp. 961–1004). Lawrence Erlbaum Associates, Inc.

Kamhi, A. G. (2014). Improving clinical practices for children with language and learning disorders. *Language, Speech, and Hearing Services in Schools*, *45*(2), 92-103.

Klee, & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, *12*(2), 251–269.

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, *12*(2), 28–41.

Klee, T., Stokes, S. F., Wong, A. M. Y., Fletcher, P., & Gavin, W. J. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*. *47*(6), 1396–1410. https://doi.org/10.1044/1092-4388(2004/104)

Klintfors, E., Lacerda, F., & Sundberg, U. (2009). Major parts-of-speech in child language– division in open and close class words. In *Fonetik 2009* (pp. 126-129). Universitetsservice US-AB.

Kong, L. D., & Fu, M. Y. (2004). Development of adverbs in children language. *Journal of Anhui Normal University (Humanities & Social Sciences) 5*, 591-596 [In Chinese: 孔令达, & 傅满义. (2004). 儿童语言中副词的发展. *安徽师范大学学报 (人文社会科学版)*, *5*, 591-596].

Kornbrot, D. (2014). Point biserial correlation. In Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118445112.stat06227

Lahey, M., & Bloom, L. (1977). Planning a first lexicon: Which words to teach first. *Journal of Speech and Hearing Disorders, 42*(3), 340-350.

Lahey, M. (1988). *Language disorders and language development*. N.Y.: Macmillan

Lau, D. K. Y., Tang, T. P. Y., Wong, C. C. Y., Yau, T. S. K., & Lam, I. H. C. (2023). Systematic language input improved productions of elaborated verb phrases of Cantonese-speaking children with language difficulties. *Clinical Linguistics & Phonetics*, 1-18.

Law, J., Garrett, Z., Nye, C., & Cochrane Developmental, Psychosocial and Learning Problems Group. (1996). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database of Systematic Reviews, 2015*(5).

Law, J., Rush, R., Schoon, I., & Parsons, S.(2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research,52*(6), 1401–1416. https://doi.org/10.1044/1092-4388(2009/08-0142)

Lee, L. L., & Canter, S. M. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*(3), 315-340.

Lee, H. T., & Wong, C. (1998). Cancorp: The Hong Kong Cantonese child language corpus. *Cahiers de Linguistique - Asie Orientale, 27*(2), 211-228.

Leonard, L. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Leonard, L. B. (2014). *Children with specific language impairment*. MIT press.

Leonard, L., Ellis Weismer, S., Miller, C., Francis, D., Tomblin, J., & Kail, R. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research, 50*, 408–428.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4*(2), 185-199.

Li, C., & Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Li, H., & Zhou, J. (2015). *Study on dinner table talk of preschool children family in Shanghai.* [Unpublished Master dissertation]. East China Normal University. Shanghai, China.

Li, H., Liang, L., & Wu, D. (2022). Predicting Chinese preschoolers' acquisition of aspect markers: A corpus-based study. *Languages*, *7*(2), 133.

Li, L., & Zhou, J. (2008)*. Metacommunication development of children aged from three to six during their collaborative pretend play*. [Unpublished Master dissertation]. The University of Hong Kong.

Li, P., & Bowerman, M. (1998). The acquisition of lexical and grammatical aspect in Chinese. *First Language*, *18*(54), 311-350.

Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. De Gruyter.

Li, Y. (2004). *The development of child language*. Wuhan: Huazhong Normal University Press.

Liang, L., Wu, D., & Li, H. (2019). Chinese preschoolers' acquisition of temporal adverbs indicating past, present, and future: A corpus-based study. *Journal of Child Language*, *46*(4), 760-784.

Lieven, E. V. M., Pine, J. M., & Barnes, H. D. (1992). Individual differences in early vocabulary development: redefining the referential-expressive distinction. *Journal of Child Language*, *19*(2), 287–310.

Liu, F. H. (2006). Dative constructions in Chinese. *Language and Linguistics*, *7*(4), 863-904.

Liu, S., Zhao, X., Li, P., Love, B. C., McRae, K., & Sloutsky, V. M. (2008). Early lexical development: A corpus-based study of three languages. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1343-1348). Austin, TX: Cognitive Science Society.

Liu, M. (2015). Tense and aspect in Mandarin Chinese. In W. S. Y. Wang & C. Sun
(Eds.), The Oxford handbook of Chinese linguistics (pp. 274-289). Oxford University
Press. https://doi.org/10.1093/oxfordhb/9780199856336.013.0070

Liu, S. (2007). Early vocabulary development in English, Mandarin, and Cantonese: A cross-
linguistic study based on CHILDES. *Master's Theses*. 705.

Long, S. (2012). Computerized Profiling'of clinical language samples and the issue of
time. *Assessing grammar: The languages of LARSP*, 29-42.

Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition.
International Journal of Corpus Linguistics, 14(1), 3-28.

Ma, W., Golinkoff, R., Hirsh-Pasek, K., Mcdonough, C., & Tardif, T. (2009). Imageability
predicts the age of acquisition of verbs in Chinese children. *Journal of Child
Language, 36*(2), 405-423.

MacLeod, A. A., Sutton, A., Trudeau, N., & Thordardottir, E. (2011). The acquisition of
consonants in Québécois French: A cross-sectional study of pre-school aged children.
International journal of speech-language pathology, 13(2), 93-109.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk*. 3rd Edition. Mahwah,
NJ: Lawrence Erlbaum Associates.

Mainela-Arnold, E., & Evans, J. L. (2014). Syntactic complexity in the spontaneous language
samples of preschool children with and without specific language impairment.
Language, Speech, and Hearing Services in Schools, 45(3), 225-237. doi: 10.1044/0161-
1461(2013/12-0061)

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency
interviews using a new measure of lexical diversity. *Language Testing, 19*(1), 85-104.

Marshall, C. R., & Van der Lely, H. K. J. (2007). Derivational morphology in children with grammatical-specific language impairment. *Clinical Linguistics & Phonetics*, *21*(2), 71-91.

Maslen, R. J. C., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech Language and Hearing Research, 47*(6), 1319-1333.

Matthews, P.H. (1974). *Morphology: An introduction to the theory of word-structure.* Cambridge, UK: Cambridge University Press.

Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar*. Routledge. https://doi.org/10.4324/9780203835012

McCauley, R. J., Fey, M. E., Gillam, R., Kamhi, A., & Leonard, L. (2016). *Treatment of Language Disorders in Children* (Second edition). Brookes Publishing.

McCormack, T., & Hoerl, C. (2008). Temporal decentering and the development of temporal concepts. *Language Learning*, *58*, 89-113.

McCormack, T., & Hoerl, C. (2017). The development of temporal concepts: Learning to locate events in time. *Timing & Time Perception*, *5*(3-4), 297-327.

McGregor, K. K. (2009). Semantics in child language disorders. In R. G. Schwartz (Ed.), *Handbook of child language disorders* (pp. 392-415). Psychology Press.

McGregor, K. K. (2020). How we fail children with developmental language disorder. *Language, speech, and hearing services in schools, 51*(4), 981-992.

McGregor, K. K., Oleson, J. J., Bahnsen, A. J., & Duff, D. (2015). Sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 58*(6), 1851-1863. doi: 10.1044/2015_JSLHR-L-14-0310

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders, 48*(3), 307-319.

Miller, J., & Chapman, R. (1985). Systematic analysis of language transcripts. *Madison, WI: Language Analysis Laboratory.*

Miller J.F. (1981) *Assessing language production in children*. Baltimore, MD: University Park Press.

Miller, J. (1991). Quantifying productive language disorders. In J. Miller (ed.), *Research on Child Language Disorders: A Decade, of Progress.* Austin: Pro-Ed.

Mok, Z., & Kipka, P. F. (2009). Lexical-semantic immaturities manifesting as grammatical disorders: evidence from a child language sample. *Clinical linguistics & phonetics*, *23*(11), 808-824.

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working Memory and Specific Language Impairment: An Update on the Relation and Perspectives on Assessment and Treatment. *American Journal of Speech-Language Pathology*, *19*(1), 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

Montgomery, J. (2005). Effects of input rate and age on the realtime lexical processing of children with specific language impairment. *International Journal of Language and Communication Disorders*, 40, 171–188

Montgomery, J., & Evans, J. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*, 269–288.

Montgomery, J., & Windsor, J. (2007). Examining the language performances of children with and without specific language impairment: Contributions of phonological short-term

memory and processing speed. *Journal of Speech, Language, and Hearing Research, 50,* 778–797.

Montgomery, J., Evans, J., & Gillam, R. (2009). Relation of auditory attention and complex sentence comprehension in children with specific language impairment: A preliminary study. *Applied Psycholinguistics, 30*, 123–151.

Moyle, M., Karasinski, C., Weismer, S., & Gorman, B. (2011). Grammatical morphology in school-age children with and without language impairment: a discriminant function analysis. *Language, Speech & Hearing Services in Schools, 42*(4), 550-560.

Na, H. L. (2017). Temporal frequency adverbs used in conjunction with the co-occurrence sequence and the expression effect. *Overseas Chinese Education*, *8*, 1093-1103 [In Chinese: 那洪雷.(2017). 时间频率副词连用共现顺序及表达效果[J]. *海外华文教育, 8*: 1093–1103].

Negen, J., & Sarnecka, B. W. (2012). Number-concept acquisition and general vocabulary development. *Child Development, 83*(6), 2019-2027.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., ... & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*(11), 1247-1257.

Oetting, J., Rice, M., & Swank, L. (1995). Quick incidental learning (QUIL) of word by school-age children with and without SLI. *Journal of Speech and Hearing Research, 38*, 434–445.

Olswang, L. B., & Bain, B. A. (1991). Intervention issues for toddlers with specific language impairments. *Topics in Language Disorders, 11*(4), 69-86.

Overton, S., & Wren, Y. (2014). Outcome measurement using naturalistic language samples: A feasibility pilot study using language transcription software and speech and language therapy assistants. *Child Language Teaching and Therapy, 30*(2), 221-229.

Owen, A. J., & Leonard, L. B. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: application of D. *Journal of Speech, Language, and Hearing Research*, *45*(5), 927–937.

Owens, R. E. (2010). *Language disorders: a functional approach to assessment and intervention* (5th ed. Ed.). Boston: Boston: Pearson/Allyn & Bacon.

Paul, R., Norbury, C., & Gosse, C. (2018). *Language disorders from infancy through adolescence : listening, speaking, reading, writing, and communicating* (Fifth edition.). Elsevier.

Paul, R. (2000). *Language disorders from infancy through adolescents* (2nd Ed.) Saint Louis, MO: Mosby-Year Book.

Pavelko, S.L. & Owens, R.E. (2017). Sampling utterances and grammatical analysis revised (SUGAR): New normative values for language sample analysis measures. *Language, Speech, and Hearing Services in Schools, 48*(3), 197-215. doi:10.1044/2017_LSHSS-17-0022.

Piaget, J. (2000). *The construction of reality in the child*. Routledge.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological bulletin*, *134*(3), 427.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, Mass: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition the acquisition of argument structure*. MIT Press.

Pomper, R., McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Ohlmann, N. (2022). Direct instruction improves word learning for children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *65*(11), 4228-4249.

Raybeck, D., & Herrmann, D. (1990). A cross-cultural examination of semantic relations. *Journal of Cross-Cultural Psychology*, *21*(4), 452-473.

Raybeck, D., & Herrmann, D. (1996). Antonymy and semantic relations: The case for a linguistic universal. *Cross-cultural research*, *30*(2), 154-183.

Rezapour, Y., Abdi, K., Rezai, H., & Aboutorabi Kashani, P. (2011). A comparison between three methods of language sampling: Freeplay, narrative speech and conversation. *Iranian Rehabilitation Journal, 9*(2), 4-9.

Rice, M. L., & Wexler, K. (1996). Limitations of mean length of utterance (MLU) in assessing language impairment. *Journal of Speech, Language, and Hearing Research, 39*(2), 234-238. doi: 10.1044/jshr.3902.234

Rice, M. L., Sell, M. A., & Hadley, P. A. (1991). Social interactions of speech, and language-impaired children. *Journal of Speech, Language, and Hearing Research, 34*(6), 1299-1307.

Rice, M. L., Smith, S. D., & Gayán, J. (2013). Convergent genetic linkage and associations to language, speech and reading measures in families of probands with Specific Language Impairment. *Journal of Neurodevelopmental Disorders, 5*(1), 28. doi: 10.1186/1866-1955-5-28

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research.*, *53*(2), 333–349.

Rice, M. L., 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.

Rinaldi, W. (2000). Pragmatic comprehension in secondary school-aged students with specific developmental language disorder. *International Journal of Language & Communication Disorders, 35*(1), 1-29.

Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J. F. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of child language, 14*(3), 433-446.

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological bulletin*, *110*(3), 426.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2010). Morphosyntactic annotation of CHILDES transcripts. *Journal of child language*, *37*(3), 705-729.

Scarborough, H., Rescorla, L., Tager-Flusberg, H., Fowler, A., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics, 12*(1), 23-46.

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics*, 11(1), 1-22.

Scarborough, H. S. (1985). *Measuring syntactic development: Index of productive syntax*. Paper presented to the Society for Research in Child Development, Toronto.

Shatz, M., Tare, M., Nguyen, S. P., & Young, T. (2010). Acquiring non-object terms: The case for time words. Journal of Cognition and Development, 11(1), 16-36. https://doi.org/10.1080/15248370903453568

Sheng, L., & McGregor, K. K. (2010). Lexical-Semantic Organization in Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 53*(1), 146–159. https://doi.org/10.1044/1092-4388(2009/08-0160)

Shi, D. (2016). Nouns and nominal phrases. In C. Huang, & D. Shi (Eds.), *A reference grammar of Chinese*. Cambridge University Press.

Shi, J., & Zhou, P. (2018). How possessive relations are mapped onto child language: A view from Mandarin Chinese. *Journal of Psycholinguistic Research*, *47*(6), 1321-1341.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson, & D. I. Slobin (Eds.), *Studies of child language development* (pp. 175-208). New York: Holt, Rinehart, & Winston.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. *The crosslinguistic study of language acquisition*, *2*, 1157-249.

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In *Typological and contextual perspectives* (pp. 219-257). Psychology Press.

Smith, A. B., & Jackins, M. (2014). Relationship between longest utterances and later MLU in late talkers. *Clinical Linguistics & Phonetics*, *28*(3), 143-152.

Smith, C. S., & Erbaugh, M. S. (2005). Temporal interpretation in Mandarin Chinese. *Linguistics*, *43*(4), 713-756. https://doi.org/10.1515/ling.2005.43.4.713

So, L. K., & Dodd, B. J. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, *22*(3), 473-495.

Stockman, I. J., & Vaughn-Cooke, F. B. (1986). Implications of semantic category research for the language assessment of nonstandard speakers. *Topics in Language Disorders, 6*(4), 15-26.

Stockman, I. J., & Vaughn-Cooke, F. B. (1992). Lexical elaboration in children's locative action expressions. *Child Development*, *63*(5), 1104-1125.

Stockman, I., & Vaughn-Cooke, F. (1982). A re-examination of research on the language of black children: The Need for a new framework. *The Journal of Education, 164*(2), 157-172.

Stockman, Ida J. (1996). The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech, and Hearing Services in Schools, 27*(4), 355-66.

Stokes, S. F., & Fletcher, P. (2000). Lexical diversity and productivity in Cantonese-speaking children with specific language impairment. *International Journal of Language & Communication Disorders*, *35*(4), 527-541.

Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. N., . . . Rai, N. K. (2012). Nouns and verbs in Chintang: children's usage and surrounding adult speech. *Journal of Child Language, 39*(2), 284-321.

Suddendorf, T. (2010). Episodic memory versus episodic foresight: Similarities and differences. *Wiley Interdisciplinary Reviews. Cognitive Science, 1*(1), 99-107.

Szagun, G. (1979). *The development of spontaneous reference to past and future: A cross-linguistic study*. Techn. Univ. Inst. f. Psychologie.

Tang, S. (2016). Aspectual system. In C. Huang & D. Shi (Eds.), *A reference grammar of Chinese* (Reference Grammars, pp. 116-142). Cambridge University Press. doi:10.1017/CBO9781139028462.006

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child Language*, *36*(5), 1115-1144.

Tardif, T. (1996). Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. *Developmental psychology*, 32(3), 492.

Thompson, P. A., Hulme, C., Nash, H. M., Gooch, D., Hayiou-Thomas, E., & Snowling, M. J. (2015). Developmental dyslexia: predicting individual risk. *Journal of Child Psychology and Psychiatry, 56*(9), 976-987.

Tillman K. A., & Barner, D. (2015). Learning the language of time: Children's acquisition of duration words. *Cognitive Psychology*, *78*(May), 57-77. https://doi.org/10.1016/j.cogpsych.2015.03.001

To, C. K., Cheung, P. S., & McLeod, S. (2013). A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones. *Journal of Speech, Language, and Hearing Research, 56*(1), 103–122

Tomasello, M. (2000). Acquiring syntax is not what you think. In D. V. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, interventions, and outcomes* (pp. 1–16). Hove, England: Psychology Press.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245-1260.

Tommerdahl, J. and Kilpatrick, C. (2013). Analyzing Reliability of Grammatical Production in Spontaneous Samples of Varying Length. *Journal of Child Language Teaching and Therapy, 29,2*, 171-183.

Tremblay, R. E. (2003). Why socialization fails. The case of chronic physical aggression. In B. B. Lahey, T. E. Moffitt, & A. Caspi (Eds.), *Causes of conduct disorder and juvenile delinquency* (pp. 182 – 224). New York: Guilford Press.

Tremblay, R. E. (2004). Decade of behavior distinguished lecture: Development of physical aggression during infancy. *Infant Mental Health Journal, 25*, 399 – 407.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*(3), 403-421.

Tse, S. K., Li, H., & Leung, S. O. (2012). Tense and temporality: how young children express time in Cantonese. *Chinese Language and Discourse*, 3(1), 35–56.

Van der Lely, H. K. (1993). Specific language impairment in children: research findings and their therapeutic implications. *International Journal of Language & Communication Disorders*, *28*(3), 247-261.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.

Wagner, L. (2018). The influence of linguistic temporal organization on children's understanding of temporal terms and concepts. *Semantics in Language Acquisition*, *24*, 45.

Watkins, R., Kelly, D., Harbers, H., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech, Language, and Hearing Research, 38*, 1349–1355.

Weber-Fox, C., & Neville, H. J. (2001). Sensitive periods differentiate processing of open- and closed-class words: An ERP study of bilinguals. *Journal of Speech, Language, and Hearing Research*, *44*(6), 1338–1353. https://doi.org/10.1044/1092-4388(2001/104)

Weist, R. (1986). Tense and aspect. In P. Fletcher & M. Garman (Eds.), *Language acquisition: Studies in first language development* (pp. 356-374). Cambridge University Press. doi:10.1017/CBO9780511620683.021

Weist, R. M. (1989). Time concepts in language and thought: Filling the Piagetian void from two to five years. *Advances in Psychology, 59*(C), 63-118.

Weist, R. M., & Buczowska, E. (1987). The emergence of temporal adverbs in child Polish. *First Language*, *7*(21), 217-229.

Weist, R. M., Wysocka, H., & Lyytinen, P. (1991). A cross-linguistic perspective on the development of temporal systems. *Journal of Child Language*, *18*(1), 67-92.

Wu, D.Y.H. (1996). Chinese childhood socialization. In M.H. Bond (Ed.), *The handbook of Chinese psychology* (pp. 143–154). London: Oxford University Press.

Wu, S. Y., Huang, R. J., & Tsai, I. F. (2019). The applicability of D, MTLD, and MATTR in Mandarin–speaking children. *Journal of communication disorders*, *77*, 71-79.

Wu, S., Zhao, J., de Villiers, J., Liu, X. L., Rolfhus, E., Sun, X., ... & Jiang, F. (2023). *Prevalence, co-occurring difficulties, and risk factors of developmental language disorder: first evidence for Mandarin-speaking children in a population-based study*. The Lancet Regional Health–Western Pacific, 34.

Wu, S.-Y. (2020). Mean length of utterance among Mandarin-speaking children with and without DLD. *Child Language Teaching and Therapy*, *36*(3), 165–179.

Xiao, H. (1999). Independence and Obedience: An Analysis of Child Socialization Values in the United States and China. *Journal of Comparative Family Studies*, *30*(4), 641–657.

Xu, Y., & Farver, J. A. M. (2009). "What makes you shy?": understanding situational elicitors of shyness in Chinese children. *International Journal of Behavioral Development*, *33*(2), 97–104.

Yew, S. G. K., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry, 54*(5), 516-524.

Yip, P. C. & Rimmington, D. (2016). *Chinese: A comprehensive grammar*. Routledge.

Zarifian, T., Modarresi, Y., Tehrani, L. G., & Kazemi, M. D. (2015). Phonetic and phonological acquisition in Persian speaking children. *On Monolingual and Bilingual Speech*, *2015*, 430.

Zhang, M., & Hudson, J. A. (2018). The development of temporal concepts: Linguistic factors and cognitive processes. *Frontiers in Psychology*, *9*, 2451.

Zhang, Y., & Zhou, J. (2020). Building a norm-referenced dataset for vocabulary assessment based on Chinese vocD and word classes. *Journal of Chinese Writing Systems*, *4*(1), 5-17.

Zhou, G. G. (2004). An investigation of the acquisition of Chinese time system. *Applied Linguistics, 4*, 33–40 [In Chinese: 周国光 (2004). 汉语时间系统习得状况的考察. *语言文字应用, 4*: 33–40].

Zhou, Y., & Xue, N. (2015). The Chinese Discourse TreeBank: A Chinese corpus annotated with discourse relations. *Language Resources and Evaluation*, *49*(2), 397-431.

Zhu, D. X. (1982). *YuFa Jiangyi*. Beijing: Commercial Press [In Chinese: 朱德熙 (1982). *语法讲义*. 北京：商务印书馆].

Zhu, M. S., Wu, J. Z., & Zhuang, X. J. (1982). Children's comprehension of several kinds of temporal words in sentences. *Acta Psychologica Sinica, 3*, 3.