The Development of Prosodic Focus-marking in Early Bilinguals' L2: A Study of Bai-Mandarin Early Bilinguals' Mandarin

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: A Bai family wearing Bai traditional cloth.

ISBN: 978-94-6093-265-6
NUR 616

The Development of Prosodic Focus-marking in Early Bilinguals' L2: A Study of Bai-Mandarin Early Bilinguals' Mandarin

De ontwikkeling van prosodische focus-markering in de tweede taal van vroeg-tweetaligen: Een studie naar het Mandarijn-Chinees van Bai-Mandarijn-Chinese vroeg-tweetaligen

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
maandag 18 december 2017
des middags te 12:45 uur

door

Zenghui Liu
geboren 15 april 1987
te Yunnan, China

Promotor:	Prof. dr. René Kager
Co-promotores:	Dr. Aoju Chen Dr. Hans Van de Velde

This dissertation was partly accomplished with financial support from the China Scholarship Council (CSC).

If I should meet thee

After long years,

How should I greet thee? —

With silence and tears.

from *When We Two Parted* by George Gordon Byron

Table of Contents

ACKNOWLEDGEMENTS

In Dutch, “geluk” can be used to indicate “happiness” or “luck”, while in Chinese, “happiness” and “luck” are expressed by two different words. There are so many connections between happiness and luck, which is the secret I thought I discovered through many years of my life. In Bai, “donkey” is a combination of “rabbit” and “loud sound”. You can interpret it as “an animal looks like a rabbit but it makes loud sounds”. Before I knew this, I never put a link between rabbit and donkey. However, after knowing this Bai word of “donkey”, I do notice that the donkey’s long ears make it looks like a rabbit. This is how language amazes me. Language builds a connection between the human wisdom and me; it brings surprise and provides a new prospect of seeing the world to me. My PhD research in linguistics is the most wonderful journey in my love of language so far. Without many people’s guidance and support, this journey would never have ended, and with pleasure.

First of all, I would like to thank my supervisors, Dr. Aoju Chen, Dr. Hans Van der Velde, and Prof. René Kager, for their firm guidance and support through my whole PhD project. Aoju opened a door to prosody and child language acquisition for me and has led me with great patience. Since the day I started, Aoju has been working side by side with me. Her passion for linguistics and her commitment to research have influenced me and will continue to have a great impact on me. I am genuinely grateful for the countless hours she spent on supervising me. Working with Aoju, I have learned how to be a good researcher. Aoju also offered me many opportunities of co-supervising BA and MA students’ thesis or internship, as well as giving guest lectures in the linguistic MA programme, from which I gained a lot of teaching and supervision experience. Together with Aoju, Hans always provides valuable suggestions and shows me possibilities. It was my great pleasure to work with Hans. I would also like to express my gratitude to René who made a lot of time for me, especially during the last two months of my writing stage. I was really impressed by his prompt response, insightful suggestions, and generous input. Without Aoju, Hans, and René, this dissertation would not have been finished.

I would like to thank Prof. dr. Lisa Cheng, Prof. dr. Marc Swerts, Prof. dr. Peter Coopmans, Prof. dr. Vincent van Heuven and Prof. Virginia Yip, for serving as my manuscript committee.

I am indebted to all the members who worked in “Get the Focus Right” project: Anna Sara Romøren, Frank Bijlsma, Joe Rodd, Martine Veenendaal, Mengru Han, Paula Cox, Saskia Verstegen, Tom Lentz, Xiaoli Dong, and Anqi Yang. It has been

a wonderful experience working with such a great group of people. Paula, thank you for the wonderful pictures you drew for the experiment with your talent and kindness. Frank, thank you for showing me how to play around with Praat scripts. Joe, thank you for all the insightful discussions we had, and for taking care of me on many occasions. Special thanks should go to Anna Sara and Anqi. I am really lucky to work closely with you two. All the lunch meetings we had and all the conferences we attended in Dublin, Cambridge, Boston, Nijmegen, and Amsterdam have been the most wonderful memory of my career so far. Thank you for your trust, support, and sharing.

I would like to express my gratitude to Prof. Huub Van den Bergh, Mattis Van den Bergh, Kirsten Schutter, Prof. Hugo Quené, Laura Boeschoten, Tom Lentz, Maartje Schulpen for helping me with statistics. Huub, thank you for keeping your office door open for my endless questions. Your knowledge, kindness, and patience have helped me to discover the beauty of statistics. I will remember your "yo- yes and no" forever together with your encouragement. I would also like to thank Prof. dr. Carols Gussenhoven, Prof. dr. Vincent van Heuven, and Prof. dr. Sun-Ah Jun for demonstrating the glamor of prosody, phonetics and laboratory phonology in master classes and summer/winter schools organized by LOT, and for discussing my study on a number of occasions.

Many thanks to the Netherlands Organization for Scientific Research (NWO), Koninklijke Nederlands Akademie van Wetenschappen (KNAW), Chinese Scholarship Council (CSC) and Utrecht Institute of Linguistics (UiL OTS) for funding my PhD research, as a part of Talent & Training China (T-T China) programme. Most of all, I would like to thank UiL OTS. Frank Wijnen, thank you for your trust in me. Maaike Schoorlemmer, thank you for your tremendous support since the day I started, and your trust in me. Thank your Yvonne van Adrichen, Martien Camphuis and other secretaries, the cleaning staff, the janitors, and the friendly faces in Trans 10. As a member of UiL OTS, I have the privilege to attend and organize EPG (Experimental phonology group) meetings and Uiltjesdagen, from which I have benefited a lot. Furthermore, thanks to the UiL OTS and the corporation among various Dutch universities and research institutes, I was able to attend ELiTU-Talks, EMLAR, LOT-summer & winter schools, talks, workshops, and conferences both in Utrecht and at other universities in the Netherlands, which was an important part of my PhD training. I am also grateful to the BA and MA students from UiL OTS, whom I taught, supervised and worked with. Thank you for being so motivated and excellent.

Data collection in China is the core of my PhD research. I am grateful to Wenju He for collecting Bai data in Xizhou County, and the participants from Jinguisi Village.

I also thank Liquan Yang and Yanzhen Zhao from Yunnan University for their feedback on Bai. Furthermore, I would like to thank all the bilingual participants from Jinhe Primary School, the children's parents for their cooperation and the Dali Mandarin speakers from Xiaguan. Special thanks should go to Shanpeng Duan and the teachers of Jinhe Primary School for their enormous support for the field work. Three months' stay in Jinguisi Village was such an unforgettable period of time for me. My dearest students, thank you for escorting me home every day, even though it only took me five minutes to walk and I knew the way quite well. My dearest colleagues in Jinhe Primary School, thank you for making an opening fire and cooking the most delicious fish from Erhai Lake to say goodbye to me. My dearest villagers, thank you for all the efforts to keep me in the village, including the discount for the carriage trips I received and the blind dates which never happened. Yuzhen He, thank you for hosting me, for giving me the huge pumpkins and Kaki fruits you grew as presents. It has been my great pleasure to live with you and feed your cow with all your detailed instructions. My stay in Xizhou County also helped me to reconnect with my friend from high school—Dan Mu. Dan, thank you for driving me around with your scooter, buying me dinner in the best restaurant in Xizhou County, and giving me tips when I struggled. Zengmao Liu, thank you for picking me up without caring that your car was totally sunk into mud and cow dung.

I would like to thank Maosi Yan, thank you for being my tutor. In my darkest moment, you told me that what matters most is what I have experienced rather than what I will or will not get.

I thank my colleagues and friends at UiL OTS. Myrthe Bergstra, Andrea Santana Covarrubias, and Quý Ngọc ĐOÀN, thank you for your warm company, I am so lucky to be your officemate. Myrthe, thank you for sharing so many important life turning points with me, for your kindness and all the help you provided in all these years. Andrea, thank you for being so warm and loving. Doan, thank you for keeping me company when I worked late. Alexia, thank you for the joyfulness you have brought. Jorik, thank you for sharing your travel stories with me. Anne-France Pinget, thank you for sharing your experience with me, and for all the insightful discussions. Assaf Toledo, thank you for the amazing photos. Sander Van Der Harst, thanks for showing me how to use COG scripts. Brigitta Keij, thank you for being so nice and helpful all the time. Carmen, thank you for making lemon tea when I was sick, for the cake and Tortilla you made, for the hugs and comfort you brought to me. I owe many thanks to all the lovely faces in Trans, they are Nadya Goldberg, Eva Poortman, Marko Hladnik, Lotte Handriks, Dominique Blok, Marjolein van Egmond, Hanna de Vries, Anne van Leeuwen, Desiree Capel, Sandrien van Ommen, Jet Hoek, Ileana Grama, Pernelle Lorette, Liv Persson, Marko Simonovic, Stavroula Alexandropoulou, Bert Le Bruyn, Anna Volkova, Mirjam Hachem, Franca

Wesseling, Hannah de Mulder, Hayo Terband, Lisa Bylinina, Carolien van den Hazelkamp, Choonkyu Lee, Hans Rutker Bosker. I would also like to give my special thanks to Rachida Ganga, thank you for teaching me EEG and answering my questions. Thanks for making EEG so easy and fun for me, and for conducting experiments with me until the late night in the lab.

I also thank the colleagues and students from Chinese College Nederland, Jane Zhang, Huishuang Lin, Xiaoli Dong, Liqiao Ye, Shuangshuang Hu, Siyang Kong, and Yu Zhang, thank you for your support and I have learned a lot from all of you.

A lovely group of my Chinese PhD peers has played an important role not only in my professional life but also in my personal life. Fang Li, Jingwei Zhang and Ao Chen, thank you for all the support; without you, I would never have a nice and cozy apartment in Laan Van Vollenhove. Ao, thank you for providing accommodation and company; it was such a wonderful time to live with you in Nieuwegein and Beijing. Yipu Wei, thank you for your insightful views of the world, for all the wonderful time we had in Stockholm, Kraków, Budapest, and on my bed, of course. Shuangshuang Hu, my office buddy, thank you for all the jokes and beautiful souvenirs you brought to me. Shanshan Fan, thank you for making me truly understand what standard deviation is. Thank you for sharing all the valuable information. Mengru Han, thank you for sending me photos of food in the late evening. Liquan Liu, thank you for all the inspiration and surprise you always can provide. I thank Yuning Sun, Diqian Liu, Yan Xia, Luying Hou, Yuan Xie, Bei Peng, Chou Mo, Na Hu, and Yuchen Li for participating in my experiments, and for working with me. Meiling Dai, Wenjun Yu and He Sun, it has been my great pleasure to meet you in the Netherlands.

My time in the Netherlands started with a short stay in UCU where I met my first group of friends. Jun Zhang, thank you for organizing all the wonderful dinners when we lived in UCU together, and for trusting me in doing good research. Qiong Li, thank you for the amazing trip to Spain, which remains as one of the best trips I had. Wentao Li, thank you for playing squash with me, which actually is a game of picking up balls for you. Thank you for sharing your childhood memory and future visions with me. Weidong Zhang, thank you for your comforting visit. Xiaochun Wang, thank you for all the jokes and stories. Mo Chen, Yunmin Zhou, Miao Zhang, and Zhongshu Liang, thank you for all the wonderful BBQs, dinners and city tours we enjoyed together. After moving out from UCU, I moved to Laan van Vollenhove, where I spent most of my time during my stay in the Netherlands. I cannot forget all the lovely people I met in Zeist. I thank Fang Li, Jingwei Zhang, Shuai Li, Meng Wang, Junting Zhang, Zidan Mao, Wenjin Lu, Dingyu Liu, Yuru Wang, Yifei Lang, Yong Guo, Bo Lou, Fang Liu, and Yulong Zhao. You guys' presence made Laan

van Vollenhove such a lovely place to live. Fang Li, without you, I literately would have nothing in my tiny apartment. Thank you for all the furniture and electronic devices. When I came back to the Netherlands from my fieldtrip, you already made a home for me. Thank you. Also, thank you for making the demonstration video for my experiment. Jingwei, thank you for all the useful things you gave me to improve my life quality, for all the suggestions you provided to my research and career. Shuai Li, thank you for all the help you provided when I was new to this building. Zidan and Wenjin, thank you so much for waiting for me at the finish line of my first 10 km and the hotpot you prepared to celebrate with me. Thank you for all the wonderful food you made, the chatting till the 1:00 am, swimming, shopping, story-telling, trips, and sharing. Yuru and Dingyu, thank you for all the walks in the summer evenings. Yifei, thank you for bringing me fish, they were so fresh and delicious.

I would like to express my appreciation to my paranymphs, Xin Li and Jeroen Breteler. Xin Li, thank you for your joyfulness and support. I had a great time with you when we were in Belgium for the summer school. Your thoughtfulness, kindness, warmness and generosity are the shining stars which will never fade away. Jeroen, thank you for picking up my calls at 23:30, for teasing me, and for not being angry at me when I do not have time to play Mahjong with you.

To separate my professional and personal life apart is extremely difficult, especially with all the great people I know from work. Here come the thanks to the people I knew from work, whom I simply want to be around all the time. They are Andrea Santana Covarrubias, Heidi Klockman, Anja Goldschmidt, Björn't Hart, Jolien Scholten, Erin Pretorius, Mike Olson, Maartje Schulpen, Etske Ooijevaar, and anyone I might have forgotten to mention. We roughly started PhD at the same time. It has been my greatest pleasure to grow together with them. Thanks for sharing all the success, doubts, joy, worries, plans and dreams with me. I cannot imagine a day in Trans without your guys. Heidi, my personality buddy, thank you for showing me what is genuine kindness, for taking me for walks on your favorite route, for spending time in proofreading my draft, for being excited for every single accomplishment I made, and for simply looking at me with caring. Jolien, thank you for the lunches on all the working Saturdays, for all the meaningful conversations we had, and for all the laughs we shared. Anja Goldschmidt, thank you for liking all my stupid stories, and for the big smile on your face whenever you saw me. Björn't Hart, my running buddy, thank you for running with me and cheering for me when I finished my first 10 km in Utrecht, thank you for all the pleasant visits. Andrea Santana Covarrubias, thank you for listening to my worries and telling me that future does not exist. Erin Pretorius, thank you for being my officemate in the very beginning of my PhD, for your kind comments on my writing. Thank you for being

connected with me in many ways. Helgard Pretorius, thank you for telling me that I should give my choice a chance. Mike Olson, dear Mike, thank you for your presence, which made my life much happier than you might expected. Thank you for being one of my best friends when I was new to the Netherlands, for always being there whenever I need you, for picking me up in Chicago, for driving me around in Madison and for sharing your life with me. Thank you for the safety and warmness you brought me. Maartje, dear Maartje, thank you for being Maartje, for making me believe in the goodness in human beings, for walks and running, for surprise dinners, presents, poems, countless calls and for comforting me when I cried my eyes out. Etske Ooijevaar, who helped with translating the summary of this dissertation into Dutch, always spoils me by saying yes to many of my ridiculous requests. Thank you, Etske, thank you for your friendship. Jane Kuehn, it was fabulous to know you and maintain our friendship for many years. Thank you for visiting me, I really enjoyed your stay. David Franssen, thank you for being so nice to me when I was quite new to the Netherlands. Anneke de Waard, Fernanda Engels, and Anton Poludněv, thank you for the wonderful trip in Brussels and Prague. Thank you for forgiving me for always being busy, and for making efforts to maintain our friendship.

There are some special people I would like to thank. Shan Fan, thank you for carrying me with your bike from Nieuwegein to Zeist. Jia Li, thank you for taking the initiative to talk to me. Thank you for cycling with me in the middle of the night to carry a closet from IKEA. Thank you for your friendship. Jialinhang Cai, thank you for flying to Kunming just to meet me. You always understand me and encourage me to pursue what I want. Xiao Chen, oh, my dear Xiao Chen, thank you for keeping me warm. Thank you for talking with me no matter how late or early that is for you. Thank you for loving me as being me. Yu Liu, thank you for being the most understanding person, for making a home for me whenever I visit you, for growing up with me.

亲爱的大伯、大妈、三叔、三姨、大姐、小谭哥、阿早姐、阿亚哥、阿锐哥、代崔苗阿姐、阿成、阿诺、阿广、外婆、老舅、二姨、腾冲三姨、腾冲三叔、四叔、四姨、五叔、五姨、秉云、哥哥、春林哥哥、美林姐姐、张楠、娇燕、雪娇，谢谢你们这么多年来的关爱和支持，在我们的大家庭长大，知道有你们做我坚强的后盾，我才能任意地前行。

亲爱的阿奶、爸爸、妈妈、哥哥、姐姐，谢谢你们教会我爱。

CHAPTER 1 Introduction

1.1 Scope of the thesis

There is a large population of children who grow up with more than one language, both globally (Grosjean, 2010; Tucker, 1998) and domestically in China.[1] However, the understanding of the language development in bilingual children is still limited, especially in the domain of prosody (Hoffmann, 2014; Paradis, 2007). This thesis concerns language development in early bilingual children. In the literature of bilingualism and language acquisition, the age of bilinguals at the time of acquisition results in the differentiation between "early bilingualism" and "late bilingualism" (Hoffmann, 2014, p. 18). The cut-off age of early bilingualism (i.e., a time point when the second language kicks in) has been under discussion for decades, ranging from three years old to puberty (Bialystok & Miller, 1999; DeKeyser, 2000; Goodluck, 1986; Guasti, 2004; Johnson & Newport, 1989; Krashen, 1973; Lakshmanan, 1995; Lenneberg, 1967; Long, 1990; McLaughlin, 2012; Penfield & Roberts, 2014; Schwartz, 2004; Unsworth, 2005). In the present study, following the age criteria set in Unsworth (2005), early bilinguals are denoted as learners whose acquisition of the second language (hereafter L2) starts between four to seven years[2]. Late bilinguals are denoted as learners whose acquisition of the second language begins at an age of eight years or older.

A central question in research on early bilingual children's L2 acquisition is whether early bilingual children's L2 acquisition resembles monolingual children's L1 acquisition (Meisel, 2004; Paradis, 2007; Unsworth, 2005). It seems that the answers depend on whether production or perception has been examined. For example,

---

[1] Mandarin is spoken by the majority of the Chinese population. However, there are 55 minority ethnic groups (one billion people) in China, and 52 of the minority ethnic groups have their own languages.

[2] "Early bilinguals" defined in the present study is an equivalent term to "child L2 learners" defined in Paradis (2007). Specifically, child L2 learners are denoted as children "who have established one language before they begin learning the other, and typically speak the L1 at home and the L2 at school" (Paradis, 2007, p. 387).

Unsworth (2005) found that English (L1)-Dutch (L2) early bilingual children's acquisition of direct object scrambling in their L2-Dutch was similar to monolingual Dutch-speaking children in comprehension, but not in production. Also, it seems that the answers depend on the specific linguistic domain under investigation, such as lexicon, syntax, and segmental phonology, that has been examined (Meisel, 2008; Paradis, 2007). In the domain of vocabulary and grammatical development, although some differences between early bilingual children's L2 and monolingual children's L1 acquisition have been found (Meisel, 2008), many studies have provided evidence that the route and rate of early bilinguals' L2 acquisition are similar to those of monolingual children's L1 (Dulay & Burt, 1973, 1974; Jia, 2003; Krashen, 1982; Paradis, 2005, 2007). For example, Paradis (2005) examined the English vocabulary knowledge of early bilingual children who had various minority languages as L1s and acquired English as an L2 in Canada and found that the early bilingual children gained twelve months developmentally in vocabulary knowledge within twelve chronological months of exposure to their L2. Thus, the overall rate of vocabulary accumulation for early bilingual children's L2 acquisition is similar to the rate for L1 acquisition. Further, Jia (2003) found that the average time of exposure for Mandarin (L1)-English (L2) early bilingual children to master the use of plural "-s" in their L2-English was similar to the average time of exposure for English-speaking monolingual children's L1 acquisition. Regarding the route of acquisition, Dulay and Burt (1973, 1974) and Krashen (1982) found that the acquisition of morphemes such as progressive "-ing" and plural "-s" were acquired earlier than past tense "-ed" and third person singular "-s" in early bilingual children's L2-English development, as in monolingual English-speaking children's L1 development. In contrast, early bilingual children's L2 phonological acquisition is not comparable to age-matched monolingual children's L1 phonological acquisition (Baker & Trofimovich, 2005; Trofimovich & Baker, 2007; Tsukada et al., 2004). For example, Tsukada, Birdsong, Mack, Sun, Bialystock, and Flege (2004) examined the release bursts in English word-final voiceless stops produced by Korean (L1)-English (L2) early bilinguals (age of acquisition: around eight years old). By comparing the bilingual children's L2-English production to the age-

matched monolingual English-speaking children's production, they found that early bilinguals were not native-like[3]. This thesis concerns prosodic development in early bilinguals' L2 in general and the acquisition of an important function of prosody, i.e., to highlight new information or to mark focus in a sentence, in particular.

Recent years have seen increasing interest in the developmental trajectory of adult-like use of prosody for encoding focus (i.e., prosodic focus-marking) in monolingual children (Chen, 2011 on Dutch-speaking children; Yang & Chen, 2017 on Mandarin-speaking children; Romøren, 2016 on Swedish- and Dutch-speaking children; Yang, 2017 on Korean-speaking children; Arnhold, Chen, & Järvikivi, 2016 on Finnish-speaking children; Müller, Höhle, Schmitz, & Weissenborn, 2006, and Grünloh, Lieven, & Tomasello, 2015 on German-speaking children; Grigos & Patel 2010 on English-speaking children). However, research on prosodic focus-marking in a bilingual context has primarily focused on bilingual adults' competence rather than on the developmental aspect (Barnes & Michnowicz, 2015; Bullock, 2009; Colantoni, 2011; Colantoni & Gurlekian, 2004; Grosser, 1997; Gut & Pillai, 2014; O'Rourke, 2005, 2012; Swerts & Zerbian, 2010; Van Rijswijk & Muntendam, 2012; Van Rijswijk, Muntendam, & Dijkstra, 2017; Zerbian, 2013). It is still not clear how early bilingual children acquire prosodic focus-marking in their L2. More specifically, questions such as whether early bilingual children's L2 acquisition shows the same rate and route of acquisition as monolingual children's L1 acquisition, and how L1 and L2-input shape the early bilingual children's acquisition of prosodic focus-marking remain to be investigated. To answer these questions, this thesis examines the developmental trajectory and ultimate attainment of prosodic focus-marking in Bai (L1)-Mandarin (L2) early bilingual's Mandarin.

Bai (Sino-Tibetan language family, Bai genus) is a tone language spoken by the Bai ethnic group (i.e., Bai people, Báizú 白族) mainly in Dali Bai Autonomous

[3] Although a comparison between age-matched monolinguals and bilinguals has typically been made in previous studies of early bilingual children's L2 acquisition, it can be problematic given the different amounts of exposure monolinguals and bilinguals receive (De Houwer, Bornstein, & Putnick, 2014; Oller, Pearson, & Cobo-Lewis, 2007).

Prefecture (*Dàlǐ báizú zìzhìzhōu* 大理白族自治州), Yunnan (*Yúnnán* 云南), southwest of China. Mandarin (Sino-Tibetan language family, Chinese genus) is a tone language spoken by Han/Chinese (Hànzú 汉族) people in the vast area of China[4]. Standard Mandarin, also termed Putonghua (Pǔtōnghuà 普通话), is the standard variety of Mandarin and the official national language of the People's Republic of China. Standard Mandarin is "widely used in the domains of science and technology, and thus has enjoyed the legal status of 'a national lingua franca' in China and is the most powerful language in the country" (Li, 2015, p. 191). Standard Mandarin is the target of Bai-Mandarin early bilinguals' L2 acquisition. In Dali Bai Autonomous Prefecture where Bai-Mandarin early bilinguals grow up, Dali Mandarin (*Dàlǐ fāngyán* 大理方言), a regional variety of Mandarin, is also widely spoken along with Bai and Standard Mandarin. Dali Mandarin is mainly spoken by Han people in the Dali area, especially in Dali City and its sound system is similar to Standard Mandarin (Lǐ, 2009; Wú & Zhāng, 1988). All Bai speakers are bilinguals who can speak Dali Mandarin and/or Standard Mandarin as their second language. However, there are speakers of Dali Mandarin who have little knowledge of Bai and live in the urban area of Dali City (*Xiàguān* 下关). Although the promotion of Standard Mandarin started from the 1950s in China, Bai and Mandarin have been in long-lasting and intensive contact for centuries in Dali. Hereafter, we use Standard Mandarin to refer to Putonghua.

Bai-Mandarin bilinguals grow up in a bilingual language environment in which Bai is mainly used at home between family members, Standard Mandarin is mainly used at school for education purposes and receptively accessed via mass media, and Dali Mandarin is spoken as an ambient language by Han people in bilinguals' surroundings (see more details in Section 1.4). Besides the use of prosody for encoding focus prosodically in Standard Mandarin (Xu, 1999; Yang, 2017), little is

[4] The genetic affiliation of Bai has been a subject of much debate (Hefright, 2011; F. Wang, 2004; Zhào & Xú, 1996). As the genetic affiliation is beyond the scope of the present study, we have adopted the classification of Bai and Mandarin reported in WALS Online (Dryer & Haspelmath, 2013).

known about whether and how prosody is used for encoding focus in Bai and Dali Mandarin.

To put this thesis in the right perspective, we will review the field of prosodic focus-marking by introducing the notion of focus, the linguistic realization of focus and the use of prosodic cues for marking focus in Asian tone languages with a focus on Standard Mandarin in section 1.2. In section 1.3, we present a review of previous investigations into the acquisition of prosodic focus-marking in bilinguals. An introduction of the sociolinguistic situation and linguistic (prosodic) features of Bai-Mandarin early bilinguals' L1-Bai and L2-input (Semi-Standard Mandarin and Dali Mandarin) is presented in section 1.4. Research questions and hypotheses are formulated in section 1.5. Finally, the structure of the thesis is outlined in section 1.6.

1.2 Prosodic focus-marking

1.2.1 Information structure and focus

The concept "focus" is a primitive of information structure. Information structure, also known as information packaging, communicative dimension or psychological articulation, is "a structuring of sentences by syntactic, prosodic, or morphological means that arises from the need to meet the communicative demands of a particular context or discourse" (Vallduví & Engdahl, 1996, p. 460). The importance of information structure to communication has long been recognized (Chafe, 1976). As there exist numerous theoretical studies of information structure with subtle to profound differences in how a host of concepts are defined (Krifka, 2008; Lambrecht, 1994; Vallduví & Engdahl, 1996), we adopt a theoretical framework of information structure from Vallduví and Engdahl (1996) to demonstrate the notion of information structure and focus.

Information structure has traditionally been considered from a bipartite perspective as containing two primitives, such as focus-ground (Jackendoff, 1972), theme-rheme (Halliday, 1967b; Halliday, Matthiessen, & Matthiessen, 2004), topic-comment (Gundel, 1988; Hockett, 1958), new-given (Clark, Haviland, & Freedle, 1977), and

focus-topic (Sgall, Hajicová, & Panevová, 1986). Illustrations of using the "topic-comment" and "focus-ground" bipartite schemes are presented in example (1a) and (1b), respectively.

(1) Speaker A: What about John? What does he do?
Speaker B: John drinks BEER.
a. [$_T$ John] [$_C$ drinks BEER] (Topic-comment)
b. [$_G$ John] [$_F$ drinks BEER] (Ground-focus)

(Examples adapted from Vallduví & Engdahl, 1996, p. 4675)

In (1), Speaker A asks "What about John? What does he do?" and Speaker B responds: "John drinks beer". As demonstrated in (1a), "John" is the "reference frame" for the sentence. Thus, "John" is the *topic* of the sentence and performs the anchoring role to the previous discourse. And, "drinks beer" makes a new contribution, and is a *comment* using the "topic-comment" bipartite scheme (Vallduví & Engdahl, 1996, p. 465). Using the "focus-ground" bipartite scheme in (1b), "John" is the non-informative, known, or expected part, thus, it can be partitioned as *ground*; and "drinks beer" is the informative, newsy, dominant, or contrary-to-expectation part, thus, it can be partitioned as *focus* (Vallduví & Engdahl, 1996, p. 462). Both "topic-comment" and "ground-focus" partition the sentence "John drinks beer" from (1) in the same way. For such cases, both of the bipartite schemes "topic-comment" and "ground-focus" are sufficient and adequate to capture the information structural distinctions in utterances.

However, these bipartite schemes are inadequate to capture the information structural distinctions in some cases (Vallduví & Engdahl, 1996, p. 467). For example, neither "topic-comment" nor "ground-focus" can capture the informational distinctions presented in (2).

[5] In (1a), "T" and "C" stand for "Topic" and "Comment", respectively; and in (1b), "G" and "F" stand for "Ground" and "Focus", respectively. Same denotations are applied to example (2).

(2) Speaker A: What about John? What does he drink?
Speaker B: John drinks BEER.
a. [$_T$ John] [$_C$ drinks BEER].
b. [$_G$ John drinks] [$_F$ BEER]

(Examples adapted from Vallduví & Engdahl, 1996, p. 468)

In (2), Speaker B's response "John drinks beer" can be understood by using the "topic-comment" bipartite scheme as in (2a), which is the same as in (1a). As demonstrated in (1a), "John" is the "reference frame" for the sentence. Thus, "John" is the *topic* of the sentence and performs the anchoring role to the previous discourse. And, "drinks beer" is what makes a new contribution, which can be referred as a *comment* using the "topic-comment" bipartite scheme (Vallduví & Engdahl, 1996, p. 465). However, by using the "focus-ground" bipartite scheme as in (2b), "John drinks" is the non-informative, known, or expected part, thus can be partitioned as *ground*, while "beer" is the informative, newsy, dominant, or contrary-to-expectation part, thus can be partitioned as *focus* (Vallduví & Engdahl, 1996, p. 462). The problematic part is "drinks". As "drink" has been mentioned in the preceding question, it is known for both of the speakers and not newsy anymore. Thus, whether it is be partitioned together with "beer" as a new contribution and consequently as *comment* in "topic-comment" bipartite scheme is questionable.

Vallduví (1990) proposed a set of tripartite primitives integrating all the bipartite primitives within one single schema: a trinominal articulation divides sentences into a *focus* and a *ground* (as in the "ground-focus" partition), while the latter is further composed of a *link* and a *tail*. *Focus* is defined as the only contribution that a sentence makes to the knowledge state of the hearer at the time of utterance; *link* is defined as the locus of update in the hearer's knowledge state, comparable to the notion of topic; and *tail* refers to the encompassing unfocused elements that are not link-like and that specifies how the update should be effected. This tripartite division not only works well with cases that can be analyzed by bipartite schemes proposed in the previous literature but also can partition difficult cases which cannot be

adequately analyzed by bipartite schemes. For example, (1) and (2) can be re-analyzed as (3) and (4).

(3) Speaker A: What about John? What does he do?
Speaker B: [$_G$ [$_L$ John]] [$_F$ drinks BEER]

(4) Speaker A: What about John? What does he drink?
Speaker B: [$_G$ [$_L$ John] drinks] [$_F$ BEER]

(Examples adapted from Vallduví & Engdahl, 1996, p. 468)

In (3), speaker B's response "John drinks beer", "drinks beer" in speaker B's response is the only contribution that the sentence makes to the knowledge state of the hearer at the time of utterance, that is, *focus*; "John" is the locus of update in the hearer's knowledge state, comparable to the notion of topic, which can be partitioned as *link*; *tail* is not present, thus the *ground* only contains *link* –"John". In (4), "beer" in speaker B's response is the only contribution that the sentence makes to the knowledge state of the hearer at the time of utterance, that is, *focus*; "John" is the locus of update in the hearer's knowledge state, comparable to the notion of topic, which can be partitioned as *link*; the *ground* contains *link* –"John" and *tail* – "drinks".

In the present study, we use "focus" to refer to the new information to the receiver in a sentence, in line with the notion of "focus" in Vallduví and Engdahl's (1996) tripartite analysis and similar to the definition in others' work (e.g., Gundel, 1999; Lambrecht, 1994). Focus can differ in the size and contrastiveness of the focal constituent. For example, the focal constituent can be on a single lexical word (*narrow focus*) or on a syntactic constituent larger than a word such as a verb phrase or the whole utterance (*broad focus*) (Gussenhoven, 2004). The focal constituent can contain information that serves as a correction or forms a direct contrast to what has been previously mentioned in the discourse (*contrastive focus*) (Chafe, 1976). Examples of various types of focus are presented in (5) to (9), reminiscent of the mini-dialogues produced in our experiment. The focused constituents appear in

square brackets, where BF indicates broad focus in (5), NF-i narrow focus on a sentence-initial constituent in (6), NF-m narrow focus on a sentence-medial constituent in (7), NF-f on a sentence-final constituent in (8), and CF-m indicates contrastive focus on a sentence-medial constituent in (9).

(5) Speaker A: Look! This picture is very blurry. I can't see anything clearly. What does the picture show?
Speaker B: [A BEAR THROWS A BALL.]

(BF: BROAD FOCUS)

(6) Speaker A: Look! A ball. There is also a waving arm. It looks like someone throws the ball. Who throws the ball?
Speaker B: [A BEAR] throws the ball.

(NF-I: NARROW FOCUS ON THE SUBJECT)

(7) Speaker A: Look! A dog and a ball. It looks like the dog does something to the ball. What does the dog do to the ball?
Speaker B: The dog [THROWS] the ball.

(NF-M: NARROW FOCUS ON THE VERB)

(8) Speaker A: Look! A rabbit, it waves its arm. It looks like the rabbit throws something. What does the rabbit throw?
Speaker B: The rabbit throws [A BALL].

(NF-F: NARROW FOCUS ON THE OBJECT)

(9) Speaker A: Look! A cat and a ball. It looks like the cat does something to the ball. I guess the cat cuts the ball.
Speaker B: The cat [THROWS] the ball.

(CF-M: CONTRASTIVE FOCUS ON THE VERB)

1.2.2 Linguistic realization of focus

Focus can be realized by different linguistic strategies, including syntactic, morphological, and prosodic strategies (Chen, 2012; Gussenhoven, 2007). Regarding the use of syntactic strategies for encoding narrow focus, the use of word order is observed in many languages and is also called "*positional focus-marking strategy*". Narrow focus can be expressed by putting the focused constituent in a particular syntactic position (e.g., the focal constituent is put immediately before the verb in Wolof), or by using a specific syntactic structure (e.g., "it" cleft in English). Further, the use of a focus particle to mark narrow focus syntactically is also found in some languages (e.g., "ga" and "wah" focus particles in Japanese and Sundanese, respectively). Regarding the use of morphological strategies for encoding narrow focus, the use of affixation to mark focus is commonly found (e.g., the verbal prefixes and suffixes used in Wolof and Irish for expressing focus).

Focus types that differed in the size and contrastivenees of the focal constituent can also be distinguished by different linguistic strategies (Gussenhoven, 2007). For example, in Efik, a contrastive focus particle is needed for expressing contrastive focus (De Jong, 1980; Gussenhoven, 1983), but not for expressing narrow focus; in Navajo, a neutral negative "doo…da" is used to express narrow focus, while "hanii" is used to express contrastive focus (Schauber, 1978). In Standard Mandarin, narrow focus is distinguished from contrastive focus in the magnitude of the lengthened duration and expanded pitch span of the focal constituent in read speech (Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015).

Besides the use of syntactic and morphological strategies for encoding focus, the use of prosodic strategies, including both phonological and phonetic use of prosody for encoding focus, is widely observed across languages (Chen, 2009; Jun & Lee, 1998; Romøren & Chen, 2015; Xu, 1999; Yang & Chen, 2017). Phonetic means refer to making gradient variations in prosodic properties, such as pitch and duration, without making phonological changes to realize prosodic prominence and to mark focus. For example, in Standard Mandarin, the pitch span and duration of the focal constituent is extended and lengthened for encoding focus, while the tonal identity

of the focal constituent is maintained (Chen & Braun, 2006; Ouyang & Kaiser, 2015; Shih, 1988; Xu, 1999; Yang, 2017). Phonological means refer to making discrete changes to realize prosodic prominence and to mark focus (Chen, 2009, 2012; Yang, 2017). For example, in English, a pitch accent can be assigned on a constituent which is in focus in a sentence (Selkirk, 1986, 1995); or all the pitch accents of the constituents which follow the focal constituent can be left out, that is, speakers of English "de-accenting" the non-focal constituents in order to make the focal constituent more prominent (Beckman, 1996; Beckman & Pierrehumbert, 1986). In Swedish, a prominence-marking H (high) tone is assigned to the end of the focal constituent for encoding focus, which leads to the two lexical pitch accents changes from HL* to HL*H and from H*L to H*LH. In Bengali, narrow focus is expressed by a different pitch accent in comparison to broad focus (Gussenhoven, 2007). In Korean, focus initiates an accentual phrase (AP), and tends to, but does not always, include the following words in the same AP (Jun & Lee, 1998). Unlike languages such as Standard Mandarin, in which phonetic means are exclusively used for realizing focus prosodically, some languages can primarily use phonological means for encoding focus, while also making use of phonetic means, such as English and Dutch. For example, in Dutch, although the same H*L pitch accent is assigned to the focal and non-focal constituent in the sentence-initial position, the duration and pitch minimum of the focal constituent is longer and lower than the non-focal counterpart (Chen, 2009).

The different linguistic strategies for realizing focus are not mutually exclusive. For example, the use of a syntactic strategy is combined with a prosodic strategy in Lekeitio Basque (Frota, 2002; Gussenhoven, 2007); the use of a morphological strategy is combined with a syntactic strategy in Irish (Gussenhoven, 2007); and the use of a syntactic strategy is combined with a prosodic strategy in English (Wells, 1986). Furthermore, some languages do not express focus in its prosody, such as Ambonese Malay (Maskikit-Essed & Gussenhoven, 2016), Northern Sotho (Zerbian, 2007), and Yucatec Maya (Gussenhoven & Teeuw, 2008; Kügler & Skopeteas, 2007). It is also worth mentioning that focus can be expressed by means of a

deviation from unmarked uses of prosody in focus-marking such as pitch raising or durational lengthening. For example, Kügler and Genzel (2012) found that focus is prosodically expressed by lowering the pitch register in Akan, a tone language that belongs to the Kwa branch of the Niger-Congo family, and is spoken in Ghana. Similar findings are also reported in Estonian (Asu & Nolan, 2007).

1.2.3 Prosodic focus-marking in Asian tone languages

As mentioned in section 1.2.2, focus is realized prosodically in both tone languages (e.g., Standard Mandarin, Vietnamese, and Cantonese) and non-tone languages (e.g., English, Dutch, German, Finnish, and Deang). Since this thesis concerns only Asian tone languages and language varieties, particularly, Bai, Standard Mandarin and Dali Mandarin, we present a review of the use of prosody for encoding focus in Asian tone languages.

In studies of Asian tone languages, the exclusive reliance on phonetic use of prosody for encoding focus is widely observed, such as Standard Mandarin (Chen & Gussenhoven, 2008; Xu, 1999; Yang, 2017), Vietnamese (Jannedy, 2007), Cantonese (Bauer, Cheung, Cheung, & Ng, 2001), and Yi (Wang, Wang, & Qadir, 2011). However, the exact use of prosodic cues such as pitch and duration in these languages is language-specific. On the one hand, both pitch and duration are exploited for encoding focus phonetically in some tone languages, such as Standard Mandarin (Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015; Xu, 1999; Yang, 2017) and Vietnamese (Jannedy, 2007, 2008). For example, Vietnamese is known for its complex tone system. It has six lexical tones and its lexical contrasts are marked by tonal (pitch) as well as laryngeal features (Yip, 2002). Jannedy's (2007) exploratory data indicated that focus is prosodically expressed by varying pitch and duration while word order remained constant in Vietnamese. Jannedy (2008) and Michaud and Vu-Ngoc (2004) also reported that pitch-excursion and duration are used for marking focus in Vietnamese, although the use of duration is found to be speaker-dependent in Michaud and Vu-Ngoc (2004). On the other hand, only duration is exploited for encoding focus phontically in some tone languages, such as Cantonese (Bauer et al., 2001; Wu & Xu, 2010), Yi (Wang et al., 2011), and Tsat

(Wang et al., 2012). For example, Cantonese has four static tones (55/T1, 33/T3, 21/T4, 22/T6) and two dynamic tones (25/T2, 23/T5).[6] Bauer, K. Cheung, P. Cheung, and Ng (2001) examined the acoustic correlates of focus in the mid-level tone of Cantonese. They found that duration is the most significant acoustic correlate of focus, while pitch is marginal in Cantonese. Wu and Xu (2010) expanded the investigation of the prosodic focus-marking in Cantonese to all lexical tones. They examined the effect of focus on four prosodic features, namely mean pitch, pitch span, duration, and intensity. Their results are consistent with those of Bauer et al. (2001). They revealed that duration and intensity are the two main prosodic cues used to mark focus in Cantonese. More specifically, their results showed that the focal constituents have significantly longer duration than pre-focal and post-focal constituents in all instances. By contrast, mean pitch variation does not play a significant role in signaling focus. They concluded that Cantonese speakers do not seem to manipulate pitch as a mean to convey focus, especially for the static tones (Tones 1, 3, 4 and 6), but might expand the pitch span for the two dynamic tones (Tones 2 and 5), in which case it is likely to be a secondary effect of increased duration. Their data suggested that speakers of Cantonese tend not to vary the pitch of the lexical tones for the purpose of focus-marking, but keep them relatively unchanged as in non-focus conditions. These results are consistent with Man (2002) and Gu and Lee (2007) in the minimal effect of focus on the tones in Cantonese. A study on Yi, a Tibetan-Burman tone language with four lexical tones, also reported the absence of using pitch for encoding focus (Wang et al., 2011).

This brief review shows that which prosodic cues are used to what extent for encoding focus can vary from language to language, and that such language-specific use of pitch or duration for encoding focus is not related to the total number of the lexical tones in a language. Specifically, Yi and Standard Mandarin are similar in terms of having the same number of lexical tones (i.e., four), while Standard Mandarin differs from Yi in that it exploits both pitch and duration for encoding

[6] Here a 5-point scale represents the tone contours in the first part of the notation, where 1 represents the lowest relative pitch level and 5 the highest. For example, T1 is represented as 55/T1 in this context.

focus (Wang et al., 2011; Xu, 1999; Yang, 2017). Similarly, although both Cantonese and Vietnamese have a relatively large number of lexical tones (i.e., six), Vietnamese uses both pitch and duration for encoding focus and Cantonese only uses duration for the same purpose (Jannedy, 2007, 2008; Wu & Xu, 2010). Furthermore, it seems that many tone languages spoken in the south of China only exploit duration for encoding focus, such as Yi, Tsat, and Cantonese. In addition, prosodic focus-marking in less prestigious languages (e.g., Yi, Tsat, and Taiwanese) is not prone to the influence of the more prestigious language (e.g., Standard Mandarin). As we mentioned earlier, unlike Standard Mandarin, only duration is used for encoding focus in Yi (spoken in Yunnan, China) and Tsat (spoken in Sanya, Hainan, China by the Utsuls) which have been in long-lasting and intensive contact with Mandarin (Wang et al., 2012, 2011). Studies of Taiwanese (S. Chen, Wang, & Xu, 2009; Xu, Chen, & Wang, 2012) also showed that Taiwanese has not been influenced by Mandarin in the manner of prosodic focus-marking, regardless of the intensive contact with Mandarin over 60 years.[7]

There are, however, some methodological concerns in the early investigations of prosodic focus-marking in Asian tone languages that may undermine the generalizability of the findings. First, early investigations mainly used a read-aloud task to elicit utterances in varied focus conditions (Wang et al., 2012, 2011; Wu & Xu, 2010; Xu, 1999). For example, a simple read-aloud task was used in the studies of Tsat and Yi. In Wang et al.'s (2011) examination of Tsat, a sentence with the structure "It is not Daddy. Mom asked the younger sister to go out and buy some areca catechu" was used to elicit the word "mom" under focus. The word "mom" was also highlighted for participants to read. A native speaker of the target language was assigned to help with recording. This native speaker would ask a participant to repeat a target sentence when he/she noticed any error or inappropriate intonation of the participant. For languages without writing systems, such as Yi, the speakers were asked to remember the target sentences well and produce them. Thus, it is highly

[7] However, previous studies of non-Asian tone languages revealed the influence of more prestigious language on less prestigious language in terms of prosodic focus-marking (Bullock, 2009), in line with the findings reported in the domains of lexicon and segmental phonology (Dài & Lǐ, 1992; Thomason, 2001; Weinreich, 1974).

possible that the speech elicited from such a read-aloud task is strictly controlled and hardly resembles the speech used in everyday communication. Previous studies have shown that the prosodic realization of focus differs in read speech and (semi-) spontaneous speech (Bard & Aylett, 1999; Chen & Gussenhoven, 2008; De Ruiter, 2010; O'Brien & Gut, 2010; Xu, 1999; Yang, 2017). For example, Chen and Gussenhoven (2008) showed that both pitch and duration were exploited to distinguish contrastive focus from narrow focus in read speech in Standard Mandarin, while Yang (2017) found that neither pitch nor duration was used for the same purpose in semi-spontaneous speech in Standard Mandarin. Second, a relatively small number of speakers and tokens were examined in previous studies. For example, in Jannedy's (2007) examination on Vietnamese, only two speakers were included, and three tokens were elicited from each focus condition and sentence type; in Wang et al.'s (2012) examination on Tsat, only two target sentences were repeated for three times in the elicitation. Third, previous investigations were particularly concerned with the use of pitch and duration in the comparison between different types of focus and post-focus (Shen & Xu, 2016; Xu et al., 2012, 2012). In previous analyses, focal, post-focal and pre-focal constituents in narrow focus condition were compared to their counterparts in the broad focus condition. It remains to be investigated whether pitch and duration are used to distinguish narrow focus from non-focus (post-focus and pre-focus) and narrow focus from contrastive focus. Fourth, previous studies hardly took the whole lexical tone system of the target language into consideration in the experimental design. For example, in the examination of Taiwanese, only one tone (i.e., Tone 1, high level) was included in the experimental design (Xu et al., 2012). This raises the question of whether prosodic focus-marking can differ in different tones.

1.2.4 Prosodic focus-marking in Standard Mandarin

As Standard Mandarin is important in this thesis, the use of prosodic cues for encoding focus in Standard Mandarin is reviewed in detail in the rest of this section. As mentioned in section 1.1 (p.4), Standard Mandarin, also termed as Putonghua (Pǔtōnghuà 普通话), is the standard variety of Mandarin and the official national

language of the People's Republic of China. Its phonetic system is based on the phonetic system of Beijing Mandarin, a regional variety of Mandarin spoken in the City of Beijing. Besides Beijing Mandarin, there are many regional varieties of Mandarin (e.g., Dali Mandarin) spoken in different areas of China, next to Standard Mandarin. As Li (2015) argued:

> In the language life in China, Putonghua is the most powerful dialect (Putonghua is considered as a dialect from the sociolinguistic perspective). Nationwide, the Han Chinese people are the core ethnic group in China and Mandarin, the Chinese language used by the Han, is widely used including in the domains of science and technology, and thus has enjoyed the legal status of "a national lingua franca" in China and is the most powerful language in the country (Li, 2015, p. 191).

Standard Mandarin is a tone language which has four lexical tones. The lexical tone system of Standard Mandarin is presented in Table 1. Using Chao's (Chao, 1930, 1968) 5-level tone scale (from 5-highest pitch level to 1-lowest pitch level), the tonal value of the Tone 1, 2, 3, and 4 are labelled as /55/, /35/, /214/, and /51/, respectively. Tone 1 is a high-level tone. Tone 2 is a high-rising tone. Tone 3 is realized as a low-dipping tone when produced in isolation or in phrase final position, which can be labelled as /214/, while only the falling part is realized when Tone 3 precedes Tone 1, 2, and 4. Due to tone sandhi,[8] Tone 3 becomes Tone 2 before another Tone 3 (Chao, 1968; Wang & Li, 1967). Tone 4 is a high-falling tone. In terms of duration, Tone 3 is found to be the longest, Tone 4 is the shortest, while Tones 1 and 2 have intermediate durations (Xu, 1997).

[8] Tone sandhi refers to the tonal alternations caused by the juxtaposition of tones or by placing a tone in a particular prosodic or morphosyntactic context (Zhang, 2010, 2014, p. 443).

Table 1. Lexical tones in Standard Mandarin (adapted from Lin, 2007, p. 89)

Tonal category	Tonal shape	Tonal value	Examples		
			IPA	English	Chinese
Tone 1	high level	55	/ma/	mother	妈
Tone 2	high-rising	35	/ma/	hemp	麻
Tone 3	low falling (-rising)	21(-4)	/ma/	horse	马
Tone 4	high falling	51	/ma/	to scold	骂

Focus is prosodically encoded via pitch and duration in Standard Mandarin (Chen, 2010; Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015; Xu, 1999; Yang, 2017; Yang & Chen, 2014). Regarding the use of prosodic cues for encoding narrow focus, the pitch span and duration of the focal constituent are expanded and lengthened relative to the non-focal counterpart (Yang & Chen, 2014). Further, the pitch span of the post-focal constituent is compressed relative to the non-focal counterpart in broad focus condition (Xu, 1999). Regarding the use of prosodic cues for encoding focus types that differed in the size of the focal constituent, the pitch span and duration of the focal constituent are expanded and lengthened in narrow focus, relative to its counterpart in broad focus in read speech (Xu, 1999). However, in semi-spontaneous speech, only the duration of the focal constituent is lengthened in narrow focus to distinguish from broad focus, while the pitch span is not varied (Yang, 2017; Yang & Chen, 2014). Regarding the use of prosodic cues for encoding focus types that differed in the contrastiveness of the focal constituent, the pitch span and duration of the focal constituent are expanded and lengthened in contrastive focus, relative to its counterpart in non-focus in read speech (Chen & Gussenhoven, 2008). However, in semi-spontaneous speech, neither pitch span nor duration is used to distinguish contrastive focus from narrow focus in semi-spontaneous speech (Yang, 2017; Yang & Chen, 2017).

It has been observed that many regional varieties of Mandarin make use of pitch and duration for encoding focus in a similar manner as in Standard Mandarin, such as regional varieties of Mandarin spoken in Nanchang (Gan dialect) (Wang et al., 2011), Lanzhou (Shen & Xu, 2016), Jinan, Liaocheng, Zibo (Duàn, Jiǎ, & Rǎn, 2013), Dalian, Harbin, Tianjin and Xi'an (Duan & Jia, 2014). Namely, speakers encode focus either by raising the mean pitch of the focal constituent or by expanding the pitch span relative to the non-focal counterpart. In addition, the compression of the pitch span of the post-focal constituent has also been consistently found in all abovementioned regional varieties of Mandarin. However, some differences in the use of duration between these varieties of Mandarin have been reported. For example, in Lan-Yin Mandarin (a regional variety of Mandarin spoken in Lanzhou), on-focus duration change was only present in sentence-medial words (Shen & Xu, 2016). Duàn et al. (2013) found that in regional varieties of Mandarin spoken in Jinan, Liaocheng, and Zibo, no noticeable difference in syllable duration between the focused and unfocused conditions was found. Further, Duan and Jia (2014) also did not find the systematic variations in syllable duration of focused constituents in comparison to unfocused constituents in Dalian, Harbin, Jinan, Tianjin and Xi'an Mandarin. Despite the differences in the use of duration, it is quite clear that pitch plays a major role in signaling focus in various regional varieties of Mandarin and Standard Mandarin.

However, limited research on varieties of Mandarin spoken in the vicinity of a language other than Mandarin have revealed different uses of prosody in focus-marking compared to Standard Mandarin. For example, Xu et al. (2012) investigated focus-marking in Taiwan Mandarin, compared to Standard Mandarin and Taiwanese. Taiwan Mandarin, spoken in Taiwan, has been in close contact with Taiwanese for several decades (Chen et al., 2009). It has been shown that Taiwan Mandarin is more similar to Taiwanese than to Standard Mandarin in terms of using prosodic cues for marking focus. Specifically, Taiwan Mandarin monolinguals not only expand pitch span but also lengthen the duration of the focal constituents relative to their non-focal counterparts for marking focus, as Standard Mandarin and Taiwanese speakers do. However, neither Taiwan Mandarin speakers nor Taiwanese

speakers produce the post-focal constituents with compressed pitch span, which is present in the Standard Mandarin speakers' production. It has been suggested that these differences between Taiwan Mandarin and Standard Mandarin can be attributed to Taiwan Mandarin's close contact with Taiwanese.

1.3 The acquisition of prosodic focus-marking in bilinguals

Early research on prosodic focus-marking in a bilingual context primarily focused on adults' competence (Barnes & Michnowicz, 2015; Bullock, 2009; Colantoni, 2011; Colantoni & Gurlekian, 2004; Grosser, 1997; Gut & Pillai, 2014; O'Rourke, 2005, 2012; Swerts & Zerbian, 2010; Van Rijswijk & Muntendam, 2012; Van Rijswijk et al., 2017; Zerbian, 2013). It has been suggested that the use of prosodic cues for encoding focus in L2 is quite difficult to acquire for late bilinguals (Backman, 1979; Chen, 2014; Gut & Pillai, 2014; He, Hanssen, van Heuven, & Gussenhoven, 2011; Kelm, 1987; McGory, 1997; Nava & Zubizarreta, 2008, 2009, Rasier & Hiligsmann, 2007, 2009; Swerts & Zerbian, 2010; Turco, Dimroth, & Braun, 2015; Ueyama & Jun, 1996; Zubizarreta & Nava, 2011). On the one hand, differences in late bilinguals' L1 and L2 prosodic systems can have a negative effect of L1 transfer on their L2 acquisition of prosodic focus-marking (Backman, 1979; Gut & Pillai, 2014; Kelm, 1987; Nava & Zubizarreta, 2008, 2009, Rasier & Hiligsmann, 2007, 2009; Swerts & Zerbian, 2010; Turco et al., 2015; Ueyama & Jun, 1996; Zubizarreta & Nava, 2011). On the other hand, similarities in the use of prosodic cues for encoding focus between learners' L1 and L2 do not necessarily bring positive L1 transfer and lead to successful learning (Chen, 2014; He et al., 2011; McGory, 1997). Although the difficulty of acquiring prosodic focus-marking is widely observed in late bilinguals, simultaneous bilinguals (i.e., bilinguals who grow up with two languages simultaneously from birth) are successful in exploiting prosodic cues for encoding focus in both of their languages (Wu & Chung, 2011), and heritage speakers (i.e., speakers who are raised in a home where a non-dominant language is spoken, who speak or merely understand the heritage language, and who are to some degree bilingual in dominant and the heritage language) are successful

in exploiting prosodic cues for encoding focus in their L2s (Chen, 2014; Hoot, 2012). The successful acquisition of prosodic focus-marking by simultaneous bilinguals and heritage speakers has been attributed to their early exposure to the target language (Chen, 2014; Wu & Chung, 2011).

Only a few studies have examined early bilinguals' prosodic focus-marking and provided different conclusions. For example, Huang and Jun (2011) found that Mandarin learners of English (AoA from five to seventeen years old) were largely native-like (Chen, 2014) in their use of accent placement to encode focus prosodically in their L2. Chen (2014) and Chen, Guion-Anderson, & Xu (2012; 2014) also found that young Quanzhou Southern Min-Mandarin (hereafter, QZSM-Mandarin) bilinguals who received more Standard Mandarin input and intensive training in Standard Mandarin produced the post-focal constituent with compressed pitch span, like monolingual speakers of Standard Mandarin. However, other studies showed that the acquisition of prosodic focus-marking in early bilinguals was unsuccessful. For example, Wang et al. (2012) found that Tsat-Mandarin early bilinguals did not vary pitch to encode focus in their Mandarin, unlike monolingual speakers of Standard Mandarin, although they lived in a Standard Mandarin-dominant environment, like QZSM–Mandarin bilinguals. Also, Gut and Pillai (2014) showed that Malay-English early bilingual only varied duration to encode focus in their English, unlike monolingual speakers of English. These mixed findings might be due to the differences in the L2-input and use of L2 between the early bilinguals examined in different studies. For example, Gut and Pillai (2014) suggested that the non-native-like prosodic focus-marking in Malay-English early bilinguals' L2-English might be related to the fact that the bilinguals had little contact with native American or British English speakers and usually received their input in English from Malaysians. Similarly, Chen et al. (2014) argued that the unsuccessful acquisition of prosodic focus-marking in older Southern Min-Mandarin early bilinguals could be explained by the lack of Beijing-Mandarin-like input, as "the teachers of the older generation probably had a lower Beijing Mandarin proficiency". However, none of these studies scrutinized the L2-input provided either by school teachers, or by the speakers of the non-standard varieties of

Mandarin in the bilinguals' language environment. The importance of the characteristics of L2-input has also been found in early bilinguals' L2 acquisition in other linguistic domains, such as segmental phonology (Flege, 2009; Flege & Liu, 2001; Piske, MacKay, & Flege, 2001). Together, these studies suggest that early exposure to L2, extensive use of L2, quality of L2 training and L2 input can be crucial to whether early bilinguals can obtain native-like competence in L2. It has also been found that early bilinguals' L1 plays an important role in their L2 acquisition of prosodic focus-marking. For example, Barnes and Michnowicz (2015) examined the peak alignment of broad focus in declaratives produced by Veneto-Spanish early bilinguals and found that the prosodic features from learners' L1 (Veneto Italian) were transferred to their L2 (Spanish) prosodic system, as least for some speakers, in line with early observations presented in vowel acquisition in Korean-English early bilinguals' L2 (Baker & Trofimovich, 2005), word stress patterns in Korean-English early bilinguals' L2 (Guion, 2005), and segmental perception in Catalan-Spanish early bilinguals (Sebastián-Gallés & Soto-Faraco, 1999).

Moreover, previous studies of early bilinguals' prosodic focus-marking in L2 were exclusively concerned with adult speakers' production. As mentioned in section 1.1, a central question in research on early bilingual children's L2 acquisition is whether early bilingual children's L2 acquisition resembles monolingual children's L1 acquisition (Meisel, 2004; Paradis, 2007; Unsworth, 2005). The answers seem to depend on the modality (production vs. comprehension) and specific linguistic domains (e.g., vocabulary and syntax) (Dulay & Burt, 1974; Jia, 2003; Meisel, 2008; Paradis, 2007; Tsukada et al., 2004; Unsworth, 2005). The question of whether early bilingual children's acquisition of prosodic focus-marking involves the same route and rate of acquisition as monolingual children's L1 acquisition remains to be investigated.

To sum up, the questions that await to be investigated are as follows: How do early bilingual children acquire prosodic focus-marking in their L2? Does early bilingual

children's L2 acquisition show the same rate and route of acquisition as the monolingual children's L1 acquisition? Can early bilinguals ultimately acquire the native-like ability to encode focus prosodically in their L2? How do L1 and the characteristics of L2-input (i.e., languages or varieties of languages available to children) shape early bilingual children's acquisition of prosodic focus-marking? This thesis addresses these questions by investigating how Bai-Mandarin early bilinguals acquire prosodic focus-marking in their L2-Mandarin.

1.4 Bai-Mandarin early bilinguals' L1 and L2

To better situate our study, we will provide an introduction to the sociolinguistic situation and linguistic (prosodic) features of Bai-Mandarin early bilinguals' L1-Bai and L2-input (Semi-Standard Mandarin and Dali Mandarin) in this section.

1.4.1 Bai-Mandarin early bilinguals

According to the latest national survey prior to the present study (Dàlǐ báizú zìzhì zhōu dìfāngzhì biānzhuàn wěiyuánhuì bàngōngshì, 2011), the majority of the residents in the Dali area are Bai (68%) and Han (26%) people.[9] As mentioned in section 1.1, Bai is a Sino-Tibetan tone language mainly spoken by Bai people. Standard Mandarin and Dali Mandarin (a regional variety of Mandarin spoken in Dali) are mainly spoken by Han people. Bai, Standard Mandarin, and Dali Mandarin are widely used in the Dali area (Dèng & Hé, 2012; Wú & Zhāng, 1988). All Bai speakers are bilinguals who can speak Dali Mandarin and/or Standard Mandarin as their second language. However, there are speakers of Dali Mandarin who have little knowledge of Bai and live in the urban area of Dali City. A number of Bai-Mandarin bilingual education programs were successfully implemented in the Central Bai-speaking area in Jianchuan County (Zhāng, 2012), but to the best of our knowledge, no Bai-Mandarin bilingual education program was implemented in the Southern Bai-speaking area (Xizhou County, where the present study was conducted).

[9] In the national survey, when an informant claims to be Bai, he/she is not necessarily a speaker of the Bai language.

Recently, Dèng and Hé (2012) conducted a language use and attitude study in Xizhou County (*Xīzhōu Zhèn* 喜洲镇), a Bai-Mandarin bilingual community in Dali. They investigated one hundred Bai people aged from 13 to 59 years old and found that 96% of the Bai people used Bai and local Mandarin (i.e., Dali Mandarin) in their daily life, and 60% of them could speak Standard Mandarin. 97% of the Bai people claimed that they used Bai with their family members at home, and 99% of them used Bai with their Bai peers in villages. Further, 95% of the Bai people claimed that they used Dali Mandarin with their Han peers in villages, and 98% of them used Dali Mandarin with their Han peers outside of the Bai-speaking community. However, 83% of the Bai people claimed that Standard Mandarin was their favorite language and 100% of them believed that Standard Mandarin was the most important language among Bai, Dali Mandarin, and Standard Mandarin.[10] Dèng and Hé (2012) concluded that Bai, Standard Mandarin, and Dali Mandarin were functionally complementary to each other. Specifically, Standard Mandarin and Dali Mandarin were mainly used in formal contexts and for education purposes, while Bai was used in informal contexts and for daily communication. Further, their survey also showed that the language choice pattern of these languages was quite stable. Standard Mandarin and Dali Mandarin were more prestigious than Bai due to their socio-economic and political prestige. In line with Dèng and Hé (2012), we also observed that Bai is mainly used at home in the Dali area. Specifically, we made observations and conducted interviews about the language used in Jinguisi Village (*Jīnguīsì Cūn* 金圭寺村), Xizhou County, where the present study was conducted. We can confirm that Bai is mainly used at home, especially due to the

[10] There is a discrepancy between the Bai people who could speak Standard Mandarin (60%) and the Bai people who preferred and valued Standard Mandarin (83% and 100%, respectively). According to our observation and living experience in Dali, we think such a discrepancy resulted from Bai people's language attitude toward Standard Mandarin. For example, a 59-year-old Bai speaker who is a farmer and has spent the majority of his/her life living in a Bai-speaking village might have limited formal education in Standard Mandarin or low Standard Mandarin proficiency, due to the fact that the promotion of Standard Mandarin started from the 1950s in China. Thus, he/she might be conservative and reluctant to claim that he/she can speak Standard Mandarin. However, as Standard Mandarin is necessary and important for social mobility and in other domains of life in China, it is possible that the Bai speaker shows his/her preference for Standard Mandarin, although he/she might have claimed that he/she cannot speak Standard Mandarin.

fact that most of the children are "left-behind children".[11] In the present Dali rural area, such as in Jinguisi Village, it is quite common that young people, including young parents, leave their village and go to Dali City or other cities for jobs. Therefore, seniors who mainly use Bai stay at home and take care of their grandchildren. As a result, Bai-Mandarin bilingual children grow up in a bilingual language environment in which Bai is mainly used at home with caregivers, Standard Mandarin is mainly used at school with school teachers, and Dali Mandarin is spoken as an ambient language by Han people in children's surroundings.

1.4.2 Bai-Mandarin early bilinguals' L1: Bai

The Bai language (白语), a Sino-Tibetan tone language, is spoken by more than one million people of the Bai ethnic group, in the southwest of China. It has three dialect zones: Central Bai (白语中部方言), Northern Bai (白语北部方言) and Southern Bai (白语南部方言) (Allen, 2004). We opted for Southern Bai in our research, as this variety of Bai is well studied in more detail at the segmental and lexical level in comparison with other varieties of Bai (Allen, 2004; Allen, Sū, & Yǐn, 1997; Dèng & Hé, 2012; Hé, 2015; Xú, 2008; Zhào & Xú, 1996). Hereafter, we will focus on Southern Bai and refer to it as Bai.

Bai has eight lexical tones from three tonal categories: level, rise, and fall (Allen, 2004; Allen et al., 1997; Zhào & Xú, 1996). The lexical tone system of Bai is presented in Table 2. Tone in Bai can be considered as a complex combination of pitch, phonation type, and degree of tenseness (Zhào & Xú, 1996). However, it is found that tenseness is not a significant feature in Southern Bai (Allen, 2004), so here we use pitch contour notation to denote eight lexical tones, rather than a combination of pitch and the feature of tense-lax in Table 2. Using Chao's 5-level tone scale (from 5-highest pitch level to 1-lowest pitch level), the tonal value of the three level tones are labelled as /55/, /44/, and /33/, respectively; the rising tone is labelled as /35/; and the four falling tones are labelled as /42/, /32/, /21/, and /31/,

[11] Left-behind children in China (*Líushǒu értóng* 留守儿童) is "the phenomenon of one of the parents or both leaving their children when they move to other areas. As a result, children are separated from their parents" (Zhōu & Duàn, 2005).

respectively. Little is known about the workings of tone sandhi in Bai. The basic word order of Bai is SVO (Xú, 2008; Zhào, 2009). To the best of our knowledge, it is still unknown whether focus is prosodically realized in Bai.

Table 2. Lexical tones in Southern Bai (adapted from Zhào & Xú, 1996, p. 486)

Tonal category	Tonal shape	Tonal value	Examples		
			IPA	English	Chinese
Tone 55	high-level	55	/tɕi/	plow (the first syllable)	犁头
Tone 44	mid-high-level	44	/tɕi/	leech	水蛭
Tone 33	mid-level	33	/tɕi/	pull	拉
Tone 42	high-falling	42	/tɕi/	nephew	侄
Tone 21	low-falling	21	/tɕi/	flag	旗子
Tone 32	mid-falling	32	/tɕi/	leak	渗出
Tone 31	mid-falling	31	/tɕi/	field	田
Tone 35	mid-rising	35	/tɕi/	positive (the first syllable)	积极

1.4.3 Bai-Mandarin early bilinguals' L2-input: Semi-Standard Mandarin and Dali Mandarin

Bai-Mandarin early bilinguals grow up in a relatively underdeveloped area in the southwest of China. While they have receptive access to Standard Mandarin via mass media, similar to monolingual Mandarin-speaking children, there might still be differences in the exposure to Standard and non-standard Mandarin between monolingual Mandarin-speaking children and Bai-Mandarin early bilingual children. To better understand the Mandarin-input of Bai-Mandarin early bilingual children, we present an overview of the composition of L2-input in Dali area in Table 3.

Table 3. Overview of L2-input for Bai-Mandarin early bilinguals' Mandarin acquisition in the Dali area

Input	Context		Manner		Age	
	informal	formal	direct[12]	indirect	< 6	> 6[13]
Standard Mandarin	+	-	-	+	+	+
Dali Mandarin	+	-	+	-	+	+
Semi-Standard Mandarin[14]	-	+	+	-	-	+

As shown in Table 3, Bai-Mandarin early bilingual children are exposed to standard and non-standard varieties of Mandarin in different context and manner at a different age. Specifically, they receive indirect input of Standard Mandarin before and after six years old in informal contexts, such as via mass media, as Standard Mandarin is used in broadcasting, TV program, and the Internet. They also receive direct input of Dali Mandarin from face-to-face by (near-) monolingual speakers before and after six years old in informal contexts, such as on the street,[15] as Dali Mandarin is widely used in the Dali area.[16] Further, Bai-Mandarin bilingual children start to frequently receive direct input of Semi-Standard Mandarin produced by school

[12] Here, we define "direct" input as having face-to-face communication in everyday communication.
[13] We set the cut-off point of age as six years old, as the age for Bai-Mandarin early bilinguals to receive formal Standard Mandarin education is six years old.
[14] Here, "Semi-Standard Mandarin" refers to Mandarin produced by school teachers who are Bai-Mandarin early bilinguals.
[15] Strictly speaking, there are no monolingual speakers of Dali Mandarin, as almost all of them can speak Standard Mandarin in the context of China. We use the term "(near-) monolingual" speakers of Dali Mandarin to indicate speakers who met all the following criteria: (1) acquire Dali Mandarin as the first language from birth; (2) using Dali Mandarin on a daily basis with self-estimated daily use exceeding 60%; (3) not having lived outside the Dali Mandarin-speaking community over the last ten years at the time of testing; and (4) not having actively used other languages for a long period on a daily basis.
[16] The (estimated) degree of exposure to Dali Mandarin for a Bai-Mandarin bilingual child (before the age of six) is less than 20%, which is insufficient for a bilingual child to be counted as a simultaneous bilingual of Bai and Dali Mandarin, following previous findings that a degree of exposure that is less than 20% to a language does not lead to an active use of that language (Pearson, Fernandez, Lewedeg, & Oller, 1997).

teachers who are Bai-Mandarin early bilinguals at the age of six in formal contexts, such as at school. For monolingual Mandarin-speaking children who grow up in Beijing (Yang, 2017), they receive both direct and indirect input of Standard Mandarin in formal and informal context before and after the age of six. However, for Bai-Mandarin early bilingual children, they mainly have direct input of Mandarin either in formal or informal context from speakers of Dali Mandarin and Semi-Standard Mandarin.

The sound system of Dali Mandarin, including consonants, vowels and lexical tones, is similar to Standard Mandarin (Lǐ, 2009; Wú & Zhāng, 1988). Dali Mandarin has four lexical tones, including a high-level tone (Tone 44), a mid-falling tone (Tone 31), a high-falling tone (Tone 53) and a low-dipping tone (Tone 213). The lexical tone system of Dali Mandarin resembles that of the Standard Mandarin (Lǐ, 2009; Wú & Zhāng, 1988), as illustrated in Table 4. Specifically, Tone 44 in Dali Mandarin corresponds to Tone 1 (55) in Standard Mandarin; Tone 31 in Dali Mandarin corresponds to Tone 2 (35) in Standard Mandarin; Tone 53 in Dali Mandarin corresponds to Tone 3 (214) in Standard Mandarin; and Tone 213 in Dali Mandarin corresponds to Tone 4 (53) in Standard Mandarin. To the best of our knowledge, it is still unknown whether focus is prosodically realized in Dali Mandarin. Neither is there any information on the production of tones and prosodic focus-marking in the Semi-Standard Mandarin produced by school teachers who are Bai-Mandarin early bilinguals.

Table 4. Corresponding relationship of lexical tones between Dali Mandarin and Standard Mandarin[17]

Mandarin variety	Tonal category	Tonal shape	Tonal value	Examples		
				IPA	English	Chinese
Standard	Tone 1	high level	55	/ma/	mother	妈
Dali		high level	44			
Standard	Tone 2	high-rising	35	/ma/	hemp	麻
Dali		mid-falling	31			
Mandarin	Tone 3	low falling (-rising)	21(-4)	/ma/	horse	马
Dali		high-falling	53			
Standard	Tone 4	high falling	51	/ma/	to scold	骂
Dali		low dipping	213			

1.5 Research questions and hypotheses

1.5.1 Research questions

Central to the present study is the question of how early bilinguals acquire prosodic focus-marking in their L2. More specifically, we aim to gain insight into two key issues: (1) Does early bilingual children's L2 acquisition show the same rate and route of acquisition as monolingual children's L1 acquisition? (2) How do L1 and L2-input shape the early bilingual children's acquisition of prosodic focus-marking? To this end, four specific research questions are raised and addressed in the main part of this thesis (chapters 2, 3, 4, and 5).

Research question 1: What is the role of prosody in focus-marking in Bai?

[17] The tonal shape, tonal value and examples of lexical tones in Mandarin are adapted from Lin (2007, p. 89), while the tonal shape and tonal value of lexical tones in Dali Mandarin are adapted from Wú and Zhāng (1988).

Research question 2: How is prosody used to encode focus in Dali Mandarin by (near-) monolingual speakers?

Research question 3: What is the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children's Mandarin?

Research question 4: What is the ultimate attainment of prosodic focus-marking in Bai-Mandarin early bilinguals' L2?

To address these specific research questions, we examine the use of prosody for encoding narrow focus and focus types that differed in constituent size and contrastiveness in different languages and language varieties produced by different groups of speakers. Specifically, Bai produced by (near-) monolingual speakers of Bai will be examined to answer Research question 1;[18] Dali Mandarin produced by (near-) monolingual speakers of Dali Mandarin will be examined to answer Research question 2; Mandarin produced by Bai (L1)-Mandarin (L2) early bilingual children will be examined to answer Research question 3, and Mandarin produced by Bai (L1)-Mandarin (L2) early bilingual adults will be examined to answer Research question 4. Our answers to Research question 3 and 4 and a subsequent comparison with monolingual Mandarin-speaking children's L1 (Yang & Chen, 2017) will enable us to address the similarities and differences in the rate and route of acquisition in early bilinguals' L2 and monolinguals' L1. In light of the answers to the four specific research questions, the influence of L1 and L2-input on the acquisition of prosodic focus-marking in early bilinguals' L2 will be addressed.

[18] Similar to the case of (near-) monolingual speakers of Dali Mandarin, strictly speaking, there are no monolingual speakers of Bai, as almost all of them can speak Standard Mandarin and/or Dali Mandarin in the context of China. We use the term "(near-) monolingual" speakers of Bai to indicate speakers who met all the following criteria: (1) acquire Bai as the first language from birth; (2) using Bai on a daily basis with self-estimated daily use exceeding 60%; (3) not having lived outside the Bai-speaking community over the last ten years at the time of testing; and (4) not having actively used other languages for a long period on a daily basis.

1.5.2 Hypotheses

With regard to Research question 1, considering that the use of pitch and duration for focus-marking in tone languages is not related to the number of lexical tones (Jannedy, 2007; Wang et al., 2011; Xu, 1999) and the lack of consistent evidence for the influence of the prestigious language on the less prestigious language in a language contact situation in prosodic focus-marking (Wang et al., 2012, 2011; Xu et al., 2012), we hypothesize that the manner of using duration and pitch for encoding focus in Bai (less prestigious language) is not similar to that in Standard Mandarin (prestigious language) (Hypothesis 1a). The predictions are that (near-) monolingual speakers of Bai will only lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and that they will vary neither duration nor pitch to distinguish focus types that differed in constituent size and contrastiveness in Bai, as found in other tone languages spoken in the south of China. An alternative hypothesis is that the manner of using duration and pitch for encoding focus in Bai is similar to that in Standard Mandarin, on the assumption that there is an influence of the prestigious language on the less prestigious language (Bullock, 2009) (Hypothesis 1b). The predictions are that speakers of Bai will expand the pitch span and lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and that they will vary duration and pitch to distinguish focus types that differed in constituent size and contrastiveness in Bai, as found in Standard Mandarin.

With regard to Research question 2, given that varieties of Mandarin spoken in the vicinity of a language other than Mandarin showed different uses of prosody for marking focus compared to Standard Mandarin, such as Taiwan Mandarin (Xu et al., 2012), we hypothesize that Dali Mandarin, which has been in long-lasting and intensive contact with Bai, exploits prosody for encoding focus in a similar manner as Bai (Hypothesis 2). The predictions are that (near-) monolingual speakers of Dali Mandarin will lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and that they will vary neither pitch-related

cues to encode narrow focus nor duration to encode focus types that differed in constituent size, like speakers of Bai.

With regard to Research question 3, previous studies of early bilinguals' L2 showed that early bilingual children's L2 acquisition can either resemble or diverge from monolingual children's L1 acquisition, depending on the modality (production vs. comprehension) and specific linguistic domains (Baker & Trofimovich, 2005; Dulay & Burt, 1973, 1974; Jia, 2003; Krashen, 1982; Paradis, 2005, 2007; Trofimovich & Baker, 2007; Tsukada et al., 2004). In the case of prosodic focus-marking, Yang and Chen (2017) found that monolingual Mandarin-speaking children learn to use duration for focus-marking purposes earlier than pitch-related cues. More specifically, they found that monolingual Mandarin-speaking children have mastered the use of duration to distinguish narrow focus from pre- and post-focus in all lexical tones, but they only use pitch-related cues for the same purpose in some lexical tones at the age of four to five years. At the age of seven to eight years, they can vary pitch span in an adult-like way to distinguish narrow focus from pre-focus in all lexical tones, and to distinguish narrow focus from post-focus in some lexical tones (i.e., Tone 2 and 4). At the age of ten to eleven years, their use of pitch-related cues is not yet fully adult-like in certain tones. Yang and Chen (2017) suggested that the earlier acquisition of duration relative to pitch as a focus-marking cue may be due to the limited acoustic space available for focus-related manipulation in pitch for learners of tone languages. Given that Bai-Mandarin early bilingual children's L1 and L2 are both tone languages, we hypothesize that Bai-Mandarin early bilingual children will learn the use of duration earlier than pitch-related cues for focus-marking purposes (Hypothesis 3a). Our prediction is then that Bai-Mandarin early bilingual children will lengthen the duration of the focal constituent to distinguish narrow focus from pre- and post-focus in all lexical tones, and use pitch-related cues for the same purpose only in some tones. Moreover, Yang and Chen (2017) have also found that monolingual Mandarin-speaking children exhibit adult-like use of prosody in distinguishing focus types that differed in constituent size earlier than in distinguishing narrow focus and non-focus. Specifically, monolingual Mandarin-

speaking children vary duration in an adult-like way to distinguish narrow focus from broad focus, regardless of tones, at the age of four to five years. However, their use of pitch-related cues is not adult-like. At the age of seven to eight years, they stop using pitch-related cues for differentiating narrow focus from broad focus, similar to monolingual Mandarin-speaking adults. At the age of ten to eleven years, they vary pitch span for distinguishing narrow focus from broad focus, different from monolingual Mandarin-speaking adults in semi-spontaneous speech, but similar to monolingual adult speakers in read speech (Xu, 1999). The earlier acquisition of using prosody in distinguishing focus types than in distinguishing narrow focus and non-focus might be explained by the fact that only duration is used to distinguish focus types whereas both duration and pitch-related cues are used to a similar extent in distinguishing narrow focus from non-focus in (semi-) spontaneous speech in Standard Mandarin (Yang, 2017). Given that Bai-Mandarin early bilingual children's L2 formal target is Standard Mandarin, and that early bilingual children's L2 acquisition can resemble monolingual children's L1 acquisition in some linguistic domains (Dulay & Burt, 1973, 1974; Jia, 2003; Krashen, 1982; Paradis, 2005, 2007), we hypothesize that Bai-Mandarin early bilingual children will learn the use of prosody in distinguishing focus types earlier than in distinguishing narrow focus and non-focus (Hypothesis 3b). Our prediction is then that Bai-Mandarin early bilingual children will master the use of duration to distinguish narrow focus from broad focus in all lexical tones earlier than the use of duration and pitch-related cues for distinguishing narrow focus from non-focus in all lexical tones.

With regard to Research question 4, considering that early exposure to L2, extensive use of L2 and quality of L2 training are crucial to whether early bilinguals can obtain native-like competence in L2 (Chen et al., 2014; Gut & Pillai, 2014; Huang & Jun, 2011; Wang et al., 2012), we hypothesize that Bai-Mandarin early bilingual adults can achieve native-like competence in prosodic focus-marking in Mandarin (Hypothesis 4). We will examine prosodic focus-marking in Bai-Mandarin bilingual adults who are teachers of Standard Mandarin at a primary school. These speakers have formally acquired Mandarin at an early age (i.e., six to seven years old). Informally, they have been exposed to Dali Mandarin (a regional variety of

Mandarin) in their daily life before the age of six. Further, they had intensive training in Standard Mandarin, use Standard Mandarin actively, have high Standard Mandarin proficiency and can thus represent the Bai-Mandarin early bilinguals with the highest level of proficiency in Standard Mandarin. Our predictions are that Bai-Mandarin early bilingual adults will master the use of duration and pitch-related cues to distinguish narrow focus from non-focus in all lexical tones and that they will master the use of duration to distinguish narrow focus from broad focus in all lexical tones, like monolingual speakers of Standard Mandarin. Their Mandarin also forms an important part of the Mandarin input available to Bai-Mandarin early bilingual children.

1.6 Outline of the thesis

Chapter 2 investigates prosodic focus-marking in Southern Bai with the aim to provide first-hand information of Bai-Mandarin early bilinguals' L1 (Research question 1).

Chapter 3 examines prosodic focus-marking in Dali Mandarin and serves the purpose of providing information on characteristics of L2-input (i.e., languages or varieties of languages spoken available to children) in Bai-Mandarin bilingual children's L2 acquisition (Research question 2).

Chapter 4 focuses on prosodic focus-marking in Mandarin produced by Bai-Mandarin early bilingual children from three age groups, including six- to seven-year olds, nine- to ten-year olds, and twelve- to thirteen-year olds. This chapter aims at revealing the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children's Mandarin (Research question 3).

Chapter 5 examines prosodic focus-marking in Mandarin produced by the Bai-Mandarin early bilingual adults who are teachers of Bai-Mandarin early bilingual children. By investigating this group of bilinguals, we aim at establishing the

ultimate attainment of prosodic focus-marking in Bai-Mandarin early bilinguals and providing information on the L2-input of bilingual children (Research question 4).

Chapter 6 summarizes the main results reported in Chapters 2, 3, 4, and 5, addresses the similarities and differences in the rate and route of acquisition between early bilingual children's L2 and monolingual children's L1 and influence of L1 and L2-input in detail, discusses the implications of the findings for the field of bilingualism, and makes suggestions for further research.

In order to make comparable comparisons between early bilinguals to monolinguals (i.e., Bai-Mandarin early bilinguals and Mandarin-speaking monolinguals), and between different languages and language varieties (i.e., Bai, Dali Mandarin, and Standard Mandarin), we have adopted the data elicitation method used in Yang and Chen (2017) and Yang (2017). We have made adjustments in the composition of the stimuli in the studies of Bai (Chapter 2) and Dali Mandarin (Chapter 3) in order to accommodate the lexical tone systems.

The main chapters are set up as separate manuscripts for publication. This approach has inevitably led to some overlap in the literature overviews in Chapters 2, 3, 4 and 5, and in the method sections. The advantage of this approach, however, is that each chapter is highly self-contained and can be read with little reference to the other chapters. In the interest of textual coherence, the repetition has been retained in this thesis. Details on the publishing status will be given in a footnote in each chapter separately.

CHAPTER 2 Prosodic Focus-marking in Bai[19]

Abstract

The present study examines prosodic marking of focus in Bai, a Sino-Tibetan language spoken in the southwest of China, by adopting a semi-spontaneous experimental approach. Bai subject-verb-object (hereafter SVO) sentences were elicited from (near-) monolingual speakers of Southern Bai in varied focus conditions via a picture game. The use of prosodic cues including duration, pitch span, pitch maximum, and pitch minimum for encoding focus was examined in different lexical tones. The results showed that Bai speakers increased the duration of the focused constituent to distinguish narrow focus from post-focus. However, Bai speakers did not use duration to distinguish focus types that differed in constituent size and contrastiveness. Further, they did not vary pitch for signaling focus and differentiating focus types that differed in the size or contrastiveness of the focal constituent. These results thus suggest that Bai uses prosody to mark focus, but to a lesser extent, compared to Mandarin, with which Bai has been in close contact for centuries.

Keywords: prosody, focus, Bai

[19] Preliminary results of a portion of the data from this chapter were presented at the 7th International Conference on Speech Prosody, and published in the proceedings (Liu, Chen, & Van de Velde, 2014). A version of this chapter is to be submitted in article form to an appropriate journal.

2.1 Introduction

The present study examines prosodic focus-marking in Bai. The Bai language is an SVO Sino-Tibetan tone language spoken in the southwest of China by more than one million Bai people (i.e., Báizú 白族).

The term "focus" has been associated with various notions (Sgall et al., 1986). In the current study, we use "focus" to refer to the new information to the receiver in a sentence (Gundel, 1999; Lambrecht, 1994). Focus can differ in the size and contrastiveness of focal constituents. For example, the focal constituent can be on a single lexical word (*narrow focus*) or on a syntactic constituent larger than a word such as a verb phrase or the whole utterance (*broad focus*). The focal constituent can contain information that serves as a correction or forms a direct contrast to what has been previously mentioned in the discourse (*contrastive focus*) (Chafe, 1976). Focus is prosodically encoded in many languages (Chen, 2009; Cooper, Eady, & Mueller, 1985; Gussenhoven, 2007; Hanssen, Peters, & Gussenhoven, 2008; Heldner, 2003; Jannedy, 2007; Wang et al., 2011; Xu, 1999; Xu et al., 2012).

Pitch is used to encode focus in many non-tone languages (Arnhold, 2016; Baumann, Becker, Grice, & Mücke, 2007; Chen, 2009; Cooper et al., 1985; Gussenhoven, 2004; Ladd, 1996). Previous studies have shown that pitch also plays an important role in signaling focus in some tone languages. For example, in Stockholm Swedish, a lexical pitch accent language with two contrasting lexical accents, a separate high tone is added to the lexical accent to mark focus, making pitch relevant for both lexical and post-lexical distinctions (Bruce, 1982; Romøren, 2016). In Standard Mandarin, a tone language with four lexical tones, pitch is used as a major prosodic cue to realize focus in addition to duration (Chen, 2010; Ouyang & Kaiser, 2015; Xu, 1999; Yang, 2017; Yang & Chen, 2014). Specifically, the pitch span and duration of the focal constituent are expanded and lengthened relative to the non-focal counterpart for marking focus in Standard Mandarin (Yang, 2017; Yang & Chen, 2014). Furthermore, the pitch span of the post-focal constitutent is compressed relative to the non-focal counterpart in broad focus condition in

Standard Mandarin (Xu, 1999). Regarding the use of prosodic cues for encoding focus types that differ in the size of the focal constituent, the pitch span and duration of the focal constituent are expanded and lengthened for distinguishing narrow focus from broad focus in read speech in Standard Mandarin (Xu, 1999). However, only the duration of the focal constituent in narrow focus condition is lengthened, relative to its counterpart in broad focus condition, while the pitch span is not varied in semi-spontaneous speech in Standard Mandarin (Yang, 2017; Yang & Chen, 2014). Regarding the use of prosodic cues for encoding focus types that differed in the contrastiveness of the focal constituent, neither pitch span nor duration is used in semi-spontaneous speech in Standard Mandarin. Similarly, it has been shown that in Vietnamese, as a tone language containing six lexical tones, both pitch and duration are used as prosodic cues for encoding focus (Jannedy, 2007, 2008). However, other tone languages do not use pitch to mark focus. For example, in Cantonese, a language with six lexical tones, pitch variation is not systematically modified to mark focus. According to Wu and Xu (2010), duration and intensity are the main acoustic correlates of focus in Cantonese; both are increased significantly in the on-focus words in any word location for all lexical tones. Besides, no decrease in mean pitch span is found in the post-focus words. In Yucatec Maya, a language with two lexical tones, pitch is only used at the lexical level and focus is not prosodically encoded (Gussenhoven, 2006; Gussenhoven & Teeuw, 2008; Kügler & Skopeteas, 2007). Such differences in prosodic focus-marking among tone languages suggest that which cues are used to what extent can vary from language to language and is not related to the total number of lexical tones in a language.

Furthermore, it seems that the language-specific use of prosodic cues for encoding focus is quite stable in language contact. Previous studies did not provide consistent evidence for the influence of the prestigious language on the less prestigious language in a language contact situation in prosodic focus-marking (Wang et al., 2012, 2011; Xu et al., 2012). For example, Wang et al. (2011) investigated prosodic realization of focus in six languages/dialects in China: Nanchang (a dialect of Mandarin), Tibetan, Uygur, De'ang, Yi, and Wa. The analyses on Yi, De'ang and

Wa are of particular interest to the current study. Like Bai, these three languages are native languages of ethnic minority groups who live in Yunnan Province or areas close to Yunnan Province, do not have well-spread writing systems and have been in contact with Mandarin for centuries. In addition, these languages belong to different language families (Yi – Tibetan-Burma, De'ang and Wa – Mon-Khmer). Wang et al. (2011) found that, different from Standard Mandarin, only duration was varied for encoding focus in De'ang, Yi and Wa. Similar patterns are also reported in Tsat, a language spoken in Sanya, Hainan, China by the Utsuls and has also been in close contact with Mandarin (Wang et al., 2012). Furthermore, studies of Taiwanese (Chen et al., 2009; Xu et al., 2012) also showed that Taiwanese has not been influenced by Mandarin in the manner of prosodic focus-marking, regardless of the intensive contact with Mandarin over 60 years.

To sum up, previous investigations in prosodic focus-marking among tone languages suggest that which cues are used to what extent can vary from language to language and is not related to the total number of lexical tones in a language. Further, it seems that the language-specific use of prosodic cues for encoding focus is quite stable in language contact. However, there are some methodological limitations in previous studies of prosodic focus-marking in tone languages and in language in contact situations. First, previous analyses were based on rehearsed lab speech. It is known that the prosodic realization of focus differs in read speech and semi-spontaneous speech (Bard & Aylett, 1999; Chen & Gussenhoven, 2008; De Ruiter, 2010; O'Brien & Gut, 2010; Xu, 1999; Yang, 2017). For example, Xu (1999) showed that the duration and pitch span of the focal constituent was lengthened and expanded for encoding narrow focus in read speech in Standard Mandarin, as found in semi-spontaneous speech in Standard Mandarin. However, Chen and Gussenhoven (2008) showed that both pitch and duration were exploited to distinguish contrastive focus from narrow focus in read speech in Standard Mandarin, while Yang (2017) found that neither pitch nor duration was used for the same purpose in semi-spontaneous speech in Standard Mandarin. The question arises as to how speakers of Bai realize focus prosodically in semi-spontaneous speech. Second, previous investigations were particularly interested in the use of pitch and duration in the comparison

between narrow focus and post-focus and between narrow focus and broad focus (Shen & Xu, 2016; Xu et al., 2012, 2012). It remains to be investigated whether pitch and duration are used to distinguish narrow focus from non-focus (post-focus and pre-focus) and narrow focus from contrastive focus. Third, lexical tones were not always controlled by the previous stimuli (e.g., Wang et al., 2011). This raises the question of whether prosodic focus-marking can differ in different tones.

In the present study, we have adopted a picture-matching game to elicit semi-spontaneous speech in different focus conditions from multiple speakers of Bai. We examine how pitch and duration are used to distinguish narrow focus from non-focus and distinguish different types of focus (narrow focus, broad focus, and contrastive focus). In addition, we take lexical tones in Bai into consideration in the experimental design. In light of the afore-reviewed work, we hypothesize that the manner of using duration and pitch for encoding focus in Bai is not similar to that in Standard Mandarin, on the assumption of no influence of the prestigious language on the less prestigious language. The predictions are that speakers of Bai will only lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and that they will vary neither duration nor pitch to distinguish focus types that differed in constituent size and contrastiveness in Bai, as found in other tone languages spoken in the south of China. An alternative hypothesis is that the manner of using duration and pitch for encoding focus in Bai is similar to that in Standard Mandarin, on the assumption of there is an influence of the prestigious language on the less prestigious language. The predictions are that speakers of Bai will expand the pitch span and lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and that they will vary duration and pitch to distinguish focus types that differed in constituent size and contrastiveness in Bai, as found in Standard Mandarin.

Bai is the native language of Bai-Mandarin early bilinguals. To the best of our knowledge, the prosodic realization of focus in Bai has not been studied. Therefore, the present study will provide first-hand information on prosodic focus-marking in Bai.

2.2 Bai and language contact with Mandarin

The Bai language is spoken by more than one million Bai people, mainly in Dali Bai Autonomous Prefecture (*Dàlǐ báizú zìzhìzhōu* 大理白族自治州), Yunnan Province (*Yúnnán* 云南), the People's Republic of China. The basic word order of Bai is SVO (Xú, 2008; Zhào, 2009). There are three dialectal varieties: Central Bai, Northern Bai and Southern Bai (see Figure 1). We decided to focus on Southern Bai in the current study, as Southern Bai is the most thoroughly studied variety at the segmental and lexical levels (Allen, 2004; Allen et al., 1997; Hé, 2015; Xú, 2008; Zhào & Xú, 1996). The consonantal inventory of Southern Bai includes twenty-seven phonologically distinct consonants: /p, ph, m, f, v, t, th, n, l, k, kh, ŋ, x, ɣ, tɕ, tɕh, ȵ, ɕ, j, ts, tsh, s, z, tʂ, tʂh, ʂ, ʐ/. Among these consonants, four consonants, i.e., /tʂ, tʂh, ʂ, ʐ/, are mainly used for Mandarin loans. The sound system of Southern Bai also includes eight vowels: /i, e, ɛ, ɑ, o, u, ɯ, ɤ/, twelve diphthongs: /iɛ, iɑ, iɑo, io, iou, iɯ, ui, uɛ, uɑ, uo, o, ou/. In addition, one vowel and two diphthongs are produced with rhoticity in Southern Bai: /ɚ, iɚ, uɚ/ (Zhao & Xu 1996: 480-487). There are eight lexical tones from three tonal categories in Southern Bai: level (55, 44, 33), rise (35), and fall (42, 21, 32, 31) (Allen, 2004; Zhào & Xú, 1996). The lexical tone in Bai can be considered as a complex combination of pitch, phonation type, and degree of tenseness (Allen, 2004). For example, Tone 21 is often used to transcribe words produced with a greater degree of breathiness, while 42 is used to transcribe a glottalized sound (Allen, 2004; Allen et al., 1997; Zhào & Xú, 1996). However, neither phonation nor tenseness is a distinctive tonal feature in Southern Bai (Allen, 2004; Allen et al., 1997). We thus use Chao's (1930, 1968) 5-level tone scale (from 5-high to 1-low) to denote the eight lexical tones in Southern Bai in this chapter: three level tones (55/high-level, 44/mid-high-level, 33/mid-level), one rising tone (35/mid-rising), and four falling tones (42/high-falling, 32/mid-falling, 21/low-falling, 31/mid-falling), following Allen (2004).

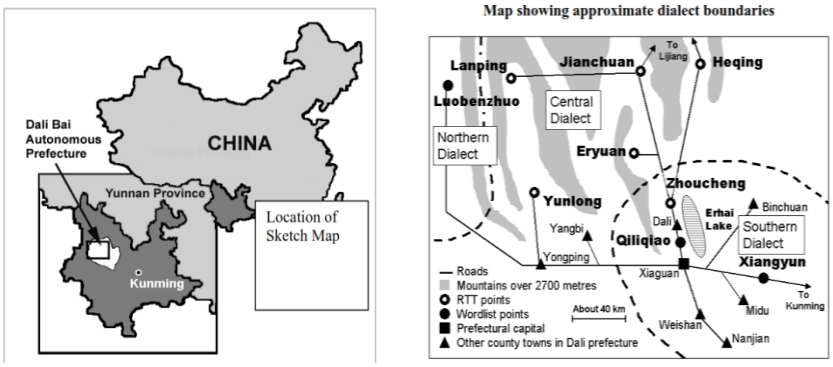

Figure 1. Speaking area of Bai varieties (adapted from Allen, 2004, p, 4)

There is some debate on the position of Bai within the Sino-Tibetan group (Allen, 2004), but this discussion is not crucial for the topic of our study. Pertinent to the current study is that Bai has been in close contact with Mandarin for centuries (Hefright, 2011). It is well documented that Bai people have been in close contact with Han people, native speakers of Mandarin, for centuries. With the formation of an integrated China in Tang Dynasty (618-907 AD), the Bai speaking area became part of China. This close contact between Bai people and Han Chinese and between their languages has continued after the foundation of the People's Republic of China in 1949. The Bai ethnic group was acknowledged as an autonomous ethnic group by the Chinese government in 1956. However, Bai is not an official language in China, not even in Dali Bai Autonomous Prefecture. Recently, Dèng and Hé (2012) have examined the language use and attitude in the Southern Bai speaking-area and found that Southern Bai and Mandarin (Standard Mandarin and Dali Mandarin) are functionally complementary to each other. In the Dali area, Standard Mandarin and Dali Mandarin (a regional variety of Mandarin) are exclusively used in education, media, and other formal contexts, while Southern Bai is used in informal contexts and for daily communication. However, Mandarin is considered as the more prestigious language due to its socio-economic and political importance.

The influence of Mandarin on Bai is readily observable in the lexicon, syntactic structure and sound system (consonant, vowel and lexical tones) (Dài & Lǐ, 1992). Regarding lexicon, Dài and Lǐ, (1992) investigated 1800 Bai words and found that over 86% of the words were Mandarin loan words. Similarly, many Bai syntactic constructions could be traced back to Mandarin's influence, such as the comparative clause, the use of connective in progressive clause, and the use of causal connective in causational clause (Dài & Lǐ, 1992). Additionally, vowels, consonants and lexical tones of Bai are heavily influenced by Mandarin. For example, Mandarin consonants like /tʂ/, /tʂh/, /ʂ/and /ʐ/, diphthongs like /ia/, /io/, and /iu/ and Tone 42 and Tone 35 are integrated into the phonological system of Bai (Dài & Lǐ, 1992). However, the influence of Mandarin on Bai sentence-level prosody has not been investigated.

2.3 Method

2.3.1 Picture-matching game

We adopted the picture-matching game used in Chen (2011) to elicit SVO sentences in Bai from native speakers in a semi-spontaneous setting for two reasons. First, there is no generally used orthography system among Bai speakers. Chinese officials have made efforts to promote an orthography system for Bai without long-term effect (Xu et al., 2007). The picture-matching game was completely based on pictures. Second, this method was also used in recent research on prosodic focus-marking in spoken Mandarin, as opposed to read speech (Yang, 2017; Yang & Chen, 2014). Using the same elicitation method enables a more reliable comparison in results of Southern Bai and Mandarin than otherwise.

In the picture-matching game, three sets of pictures were used. The experimenter who was a native speaker of Southern Bai and the participant each held a set of pictures ordered in a specific sequence; and the third set of pictures were scattered around on a table. In the experimenter's pictures (the first set), there was always some information missing, e.g., a subject, an action (verb), an object, or all three. The participant's pictures (the second set) all contained a complete event. The participant's task was to help the experimenter with sorting out pictures from the

first set (experimenter's pictures) and the third set (scattered around on the table) in matched pairs (see Figure 2).

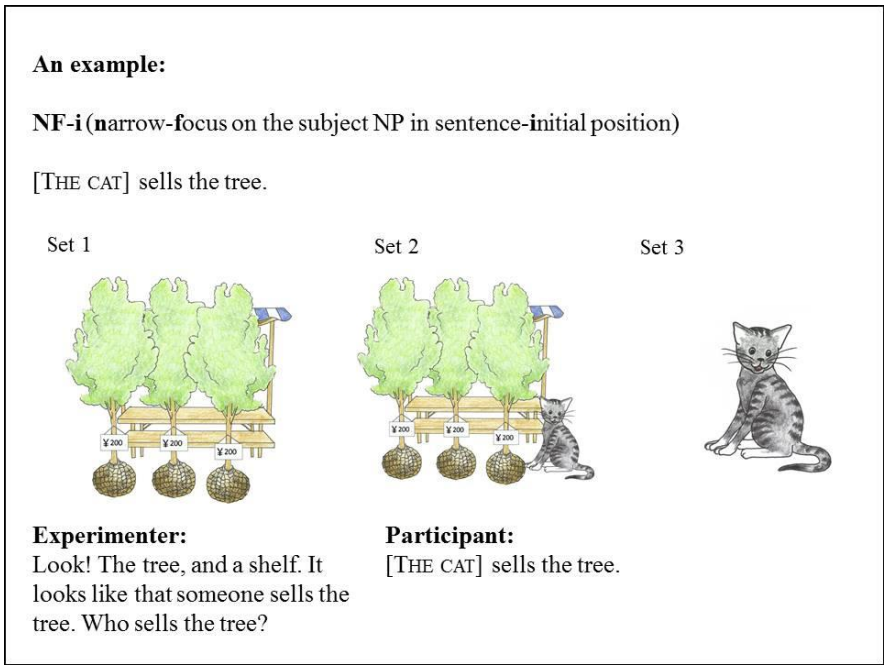

Figure 2. An example of a trial eliciting the sentence "[THE CAT] sells the tree."

An example of a trial eliciting a target sentence with **n**arrow-**f**ocus on the subject NP in sentence-**i**nitial position (the NF-i condition) is as follows: First, the experimenter takes a picture (e.g., a tree) from her own set (set 1), draws the participant's attention to the picture and establishes what the picture is by saying, e.g., "Look! The tree and a shelf. It looks like someone sells the tree." This is done to ensure that the entity in the picture is referentially given to the participant before the utterance of the question. Second, the experimenter asks a question about the picture (e.g., "Who sells the tree?"). Third, the participant takes a picture from his or her set (set 2) and inspects it. The experimenter then repeats the question, followed by an answer from the participant (e.g., "[THE CAT] sells the tree."). Lastly, the experimenter finds the picture containing the missing information in the third set (set 3) and pairs it up with

her own picture. As to the rules of the game, participants were asked to respond in full sentences and not to show their pictures to the experimenter.

To ensure consistency in the participants' word choice, the picture-matching game was preceded by a picture-naming task, which was designed to familiarize the participants with the target words and the entities in the pictures used in the game.

2.3.2 Experimental materials

The Bai experimental materials are designed by the author with help from the experimenter and a (near-) monolingual speaker of Bai. SVO sentences in five focus conditions were elicited via the picture-matching game: **n**arrow-**f**ocus on the subject NP in sentence-**i**nitial position (NF-i), **n**arrow-**f**ocus on the verb in sentence-**m**edial position (NF-m), **n**arrow-**f**ocus on the object NP in sentence-**f**inal position (NF-f), **b**road **f**ocus (BF) and **c**ontrastive-**f**ocus on the verb in sentence-**m**edial position (CF-m). The focus condition was set up by a WH-question or a statement from the experimenter, as illustrated in examples (1) to (5), where the focal constituent appears in square brackets. Each focus condition was realized in 24 SVO sentences.

(1) **NF-i**

Experimenter: Look! The tree. There is also a shelf. It looks like someone sells the tree. Who sells the tree?

Participant: [THE CAT] sells the tree.

a^{55} mi^{55} tw^{21} kɯ^{21} tsɯ^{33}.

Cat one (quantifier) sell tree.

(2) **NF-m**

Experimenter: Look! The rabbit and the tree. It looks like the rabbit does something to the tree. What does the rabbit do to the tree?

Participant: The rabbit [SELLS] the tree.

thɔ^{55} lɔ^{55} tw^{21} kɯ^{21} tsɯ^{33}.

Rabbit one (quantifier) sell tree.

(3) **NF-f**

Experimenter: Look! The bear, it stands behind a shelf. It looks like the bear sells something. What does the bear sell?

Participant: The bear sells [THE TREE]. (NF-f)

ɕo^{42} tɯ^{21} kɯ^{21} tsɯ^{33}.

Bear one (quantifier) sell tree.

(4) **BF**

Experimenter: Look! This picture is very blurry. I can't see anything clearly. What does the picture show?

Participant: [THE CAT SELLS THE TREE].

a^{55} mi^{55} tw^{21} kɯ^{21} tsɯ^{33}.

Cat one (quantifier) sell tree.

(5) **CF-m**

Experimenter: Look! The dog and the tree. It looks like the dog does something to the tree. I guess the dog wipes the tree.

Participant: The dog [SELLS] the tree.

khua^{33} tw^{21} kɯ^{21} tsɯ^{33}.

Dog one (quantifier) sell tree.

In order to keep the experiment within a feasible length, we included one lexical tone for each of the three tonal types: level tones (Tones 55, 44, and 33), falling tones (Tones 42, 31, 32, and 21) and the rising tone (Tone 35). The included lexical tones were well spread over the tonal space of Southern Bai's tone system: Tone 55 represented the level tones; Tone 21 represented the falling tones, and the rising tone: Tone 35.

In Southern Bai, a noun needs to be followed by a quantifier to form a noun phrase as a subject of a sentence, but the quantifier of the noun phrase can be omitted when the noun phrase is an object of a sentence (Zhào, 2009). Four subject-nouns were selected, which were followed by the same low falling-tone quantifier in all the target sentences. Each of the three tone types (i.e., high-level tone/Tone 55, low-falling tone/Tone 21 and rising tone/Tone 35) occurred twice in verbs. The six monosyllabic verbs and four monosyllabic object nouns formed 24 VPs, each of which appeared in each focus condition (n = 5). This resulted in 120 VPs. The four subject-noun phrases (which ended with the same quantifier "tɯ^{21}") were evenly distributed over these 120 VPs to form 120 unique SVO target sentences. These 120 target sentences were split into two lists, and each list contained all the five focus conditions realized in different sentences, and all the six representations of the tones, but only half of the VPs. This resulted in 60 sentences per list and participant. An overview of the words in SVO sentences is provided in Table 1 and original scripts glossed into IPA are provided in Appendix C.

Table 1. Overview of the words in SVO sentences, each word is presented in IPA and English

Part of speech	Gloss	Noun		Quantifier
		1st syllable	2nd syllable	3rd syllable
Subject	cat (a cat)	a^{55}	mi^{55}	tɯ^{21}
	bear (a bear)	ɕo^{42}		tɯ^{21}
	dog (a dog)	khua^{33}		tɯ^{21}
	rabbit (a rabbit)	thɔ^{55}	lɔ^{55}	tɯ^{21}
Verb (for list 1)	smell	tshu^{55}		
	wrap	pɔ^{35}		
	sell	kɯ^{21}		
Verb (for list 2)	draw	xua^{55}		
	wipe	ma^{35}		
	transport	pa^{21}		
Object	pear	ɕui^{55}	li^{55}	
	grass	tshu^{33}		
	tree	tsɯ^{33}		
	pencil	fɣ̩^{44}	a^{44}	

2.3.3 Participants and procedure

Fourteen (near-) monolingual speakers of Southern Bai (six male and eight female, age range: 22 to 46 years; mean age: 30.1 years old; SD = 9.0) participated in our study (see Appendix D). They were recruited from villages in Xizhou County, Dali Bai Autonomous Prefecture, China. All the participants had predominantly used Bai in their daily life, although they had learned Standard Mandarin and/or Dali Mandarin as their second language. The participants all met the following criteria: (1) acquire Bai as the first language from birth; (2) using Southern Bai on a daily basis with self-estimated daily use exceeding 60%; (3) not having lived outside the Southern Bai-speaking community over the last ten years at the time of testing; (4) not having actively used Mandarin or other languages for a long period on a daily

basis; and (5) having no self-reported speech and hearing impairments. They were all paid a small fee for their participation. Participants were randomly assigned to one of the two lists, and tested individually by a female experimenter, who was a native speaker of Southern Bai (age = 26). Each experiment was split into two sessions, each contained 60 trials. A break was introduced after one session to avoid loss of attention or fatigue in the participants. Each session took about 20 to 25 minutes for a participant to complete. All the test sessions were conducted in Southern Bai, in a quiet room at a villager's private home in Jinguisi Village, Xizhou County. The experiments were recorded using a portable ZOOM H1 digital recorder at a 44.1 kHz sampling rate and 16 bit accuracy. Each session was also video-recorded for future training purpose.

2.3.4 Acoustic annotation

The auditory recordings from the participants were first orthographically annotated. An answer sentence was discarded from analysis if it had one of the following features: (1) it was not uttered as a direct response to the target question; (2) it contained self-corrections or hesitations (defined as a long "em" sound produced by the participants before answering questions); (3) it deviated from the target sentence in word choice or word order; (5) it was severely overlapped with background noise coming from the farm. In total 62% of the obtained responses (n = 697) were included for further analysis.

The **verbs** were the targets of prosodic analysis. They were focal constituents in the NF-m, BF and CF-m conditions, post-focal constituents in the NF-i condition, and pre-focal constituents in the NF-f condition. The verbs were acoustically annotated by examining the waveform, wide-band spectrum, and pitch track in Praat (Boersma & Weenink, 2006) in combination with auditory impressions (Turk, Nakai, & Sugahara, 2006). Two pitch-related landmarks and two segmental landmarks were labelled in each verb: pitch maximum, pitch minimum, word onset, and word offset. An example of the annotation is presented in Figure 3.

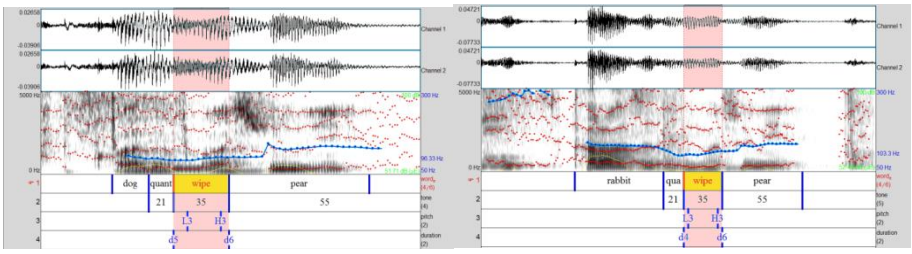

Figure 3. Pitch contour (in Hz) of verbs in NF-m vs. NF-f (NF-m = verb in sentence medial focused position, NF-f = verb preceding a focused constituent)

The tonal target of the lexical tones had been taken into consideration when the landmarks for pitch maximum and minimum were inserted (Xu, 1999; Xu & Wang, 2001; Yang & Chen, 2017). Specifically, Tone 55 is a high-level tone and appeared as a slightly rising pitch contour and occasionally as a slightly falling pitch contour sentence-medially. The pitch-maximum was labelled after the pitch-minimum in the case of a rising Tone 55 syllable but before the pitch-minimum in the case of a falling Tone 55 syllable. Tone 35 is a rising tone, but we observed that it appeared as a fall-rise contour sentence-medially, as reported for its counterpart (Tone 2) in Standard Mandarin in Xu (1997) and Yang and Chen (2017). As the falling part is largely influenced by the preceding tone, the rising part is taken to contain the tonal targets of Tone 35 (Xu, 1997). Thus, the pitch-maximum was labelled after the pitch-minimum in the rising part of Tone 35. Similarly, Tone 21 is a falling tone, but we observed that it appeared as a rise-fall contour sentence-medially, again, as reported for its counterpart (Tone 4) in Standard Mandarin in Xu (1997) and Yang and Chen (2017). As the rising part is subjected to the influence of the preceding tone, the falling part contains the tonal targets of Tone 21. Thus, the pitch-maximum was labelled before the pitch-minimum in the falling part of Tone 21.

The pitch values (in Hz) at the pitch-related landmarks, and time values (in seconds) at the segmental landmarks were automatically extracted by means of Praat scripts. Four measures were obtained from each verb: pitch maximum, pitch minimum, pitch span (i.e., the difference between the maximum pitch and the minimum pitch), and

word duration. In 115 of the usable responses (16.5%), an accurate measurement of pitch values was not possible. These responses were thus excluded from the analysis of pitch-related measurements.

2.4 Statistical analyses and results

2.4.1 Statistical analyses

In order to assess the effect of focus, we compared the measurements of the verbs between the unfocused conditions and the focused condition; that is, NF-m (verb is the focal constituent) vs. NF-i (verb is the post-focal constituent), and NF-m (verb is the focal constituent) vs. NF-f (verb is the pre-focal constituent). To investigate the effect of focus types that differed in the size of the focal constituent, we compared narrow focus on the verb (NF-m) with broad focus (BF). To investigate the effect of focus type differs in the contrastiveness of the focal constituent, we compared contrastive focus on the verb (CF-m) with non-contrastive focus on the verb (NF-m).

Statistical analyses were conducted using mixed-effects modelling in R (R Core Team, 2014) with package “lme4” (Bates, Mächler, Bolker, & Walker, 2015) and “lmerTest” (Kuznetsova, Brockhoff, & Christensen, 2013). In all models, the tone and focus conditions were included as fixed factors, while the speaker (i.e., the participants) and sentence (i.e., the answer sentences) were included as random factors. In each of the four pair-wise comparisons listed above, focus condition had two levels (the two focus conditions of interest), and tone referred to the lexical tones of the target verbs, which had three levels (i.e., “Level tone”, “Rising tone”, and “Falling tone”). Outcome variables were the duration, pitch span, pitch maximum, and pitch minimum of the verbs.

Following Magezi (2015) and Field, Miles & Field (2012), our models were constructed and evaluated in a stepwise fashion. Specifically, we started from an intercept-only model, built the subsequent models by adding one new term/factor at a time, and then systematically compared the models which differed only by the presence or absence of one term. The comparison was done using a likelihood-ratio

test, and the test statistic χ^2, degrees of freedom and *p*-value were reported for the added term. A *p*-value of less than 0.05 was considered to indicate that the added term contributed significantly to the model fit (Magezi, 2015). A detailed description of the model building is presented in Table 2.

Table 2. Model build-up procedure

Model	Added term
Model 0	only contains speaker, sentence as random intercepts
Model 1	+ tone
Model 2	+ focus condition
Model 3	+ tone : focus condition

When building the models, only the factors and interactions that significantly improved the fit of the model were retained until the best fit model was determined.[20] After the best-fit model was established,[21] we only summarized and interpreted the model with the best fit. For each analysis, we reported the results of the model comparison first, and then the parameter estimates of the best-fit model. As we are primarily interested in the effect of focus on the outcome variables, we will concentrate on the main effect of focus or interactions involving the factor of focus. However, we do not discuss in detail the main effects of the factors when the interactions involving these factors are significant.

2.4.2 Duration

2.4.2.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The effect of focus (focus vs. post-focus) on duration was tested with models specified in Table 2. The results of the fit

[20] See Bates, Kliegl, Vasishth, and Baayen (2015) for an explanation of the importance of parsimonious models.
[21] We used "anova" function to conduct the comparisons between different models (Quené & Van den Bergh, 2008). See Baayen et al., (2008) on maximum likelihood.

and model comparisons are shown in Table 3. As can be seen, the best-fit model was Model 2, which contained the main effects of tone, χ^2 (2) = 73.648, p < .001, and focus condition, χ^2 (1) = 5.006, p < .05. Parameter estimates of the best-fit model are presented in Table 4. The main effect of focus condition was such that the verbs were significantly longer in narrow focus (165.8 ms, SD = 54.5) than in post-focus (152.3 ms, SD = 47.9), regardless of tone (b = 12.460, df = 46.3, t = 2.295, p < .05), as shown in Figure 4.

Table 3. Duration, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1484.8				
1 + tone	6	−1448.0	0 vs 1	73.648	2	.000***
2 + focus condition	7	−1445.5	1 vs 2	5.006	1	.025*
3 + tone : focus condition	9	−1445.4	2 vs 3	0.123	2	.941

Table 4. Duration, narrow focus (NF-m) vs. post-focus (NF-i), parameter estimates of the best-fit model (Model 2)

	Estimate	SE	Df	t value	Pr (>|t|)
Fixed part					
Intercept	102.059	6.583	48.720	15.504	.000***
Level tone	92.559	6.671	46.640	13.875	.000***
Rising tone	56.231	6.669	46.660	8.432	.000***
Narrow focus	12.460	5.430	46.300	2.295	.026*
Random part	*name*	S^2	*SE*		
Sentence	Intercept	225.1	15.00		
Speaker	Intercept	192.5	13.87		
Residual		772.7	27.80		

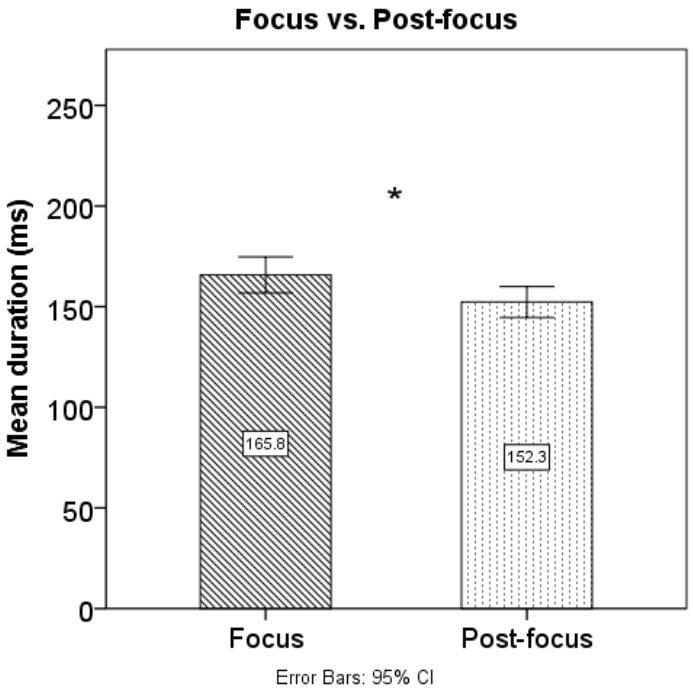

Figure 4. Mean duration (in ms) of focal constituent vs. post-focal constituent, n = 297, N = 14. Significant differences are marked with an asterisk.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 5. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (2) = 67.786, p < .001. There was thus no evidence that Bai speakers varied duration to distinguish narrow focus from pre-focus, regardless of tone.

Table 5. Duration, narrow focus (NF-m) vs. pre-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1472.2				
1 + tone	6	−1438.3	0 vs 1	67.786	2	.000***
2 + focus condition	7	−1436.5	1 vs 2	3.655	1	.056.
3 + tone : focus condition	9	−1436.0	2 vs 3	0.886	2	.642

2.4.2.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The effect of size (narrow focus vs. broad focus) on duration was tested with models specified in Table 2. The results of the fit and model comparisons are shown in Table 6. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, $\chi^2(2) = 56.328$, $p < .001$. There was thus no evidence that Bai speakers varied duration for distinguishing narrow focus from broad focus, regardless of tone.

Table 6. Duration, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1403.1				
1 + tone	6	−1375.0	0 vs 1	56.328	2	.000***
2 + focus condition	7	−1374.9	1 vs 2	0.165	1	.685
3 + tone : focus condition	9	−1374.2	2 vs 3	1.456	2	.483

2.4.2.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 7. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 65.465, $p < .001$. There was thus no evidence that Bai speakers varied duration to distinguish contrastive focus from non-contrastive focus, regardless of tone.

Table 7. Duration, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1456.8				
1 + tone	6	−1424.1	0 vs 1	65.465	2	.000***
2 + focus condition	7	−1423.6	1 vs 2	0.946	1	.331
3 + tone : focus condition	9	−1423.6	2 vs 3	0.058	2	.972

2.4.3 Pitch span

2.4.3.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 8. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 30.724, $p < .01$. There was thus no evidence that Bai speakers varied pitch span to distinguish narrow focus from post-focus, regardless of tone.

Table 8. Pitch span, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−963.55				
1 + tone	6	−948.19	0 vs 1	30.724	2	.000***
2 + focus condition	7	−947.84	1 vs 2	0.710	1	.399
3 + tone : focus condition	9	−947.09	2 vs 3	1.489	2	.475

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 9. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 32.877, $p < .001$. There was thus no evidence that Bai speakers varied pitch span to distinguish narrow focus from pre-focus, regardless of tone.

Table 9. Pitch span, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−999.17				
1 + tone	6	−982.73	0 vs 1	32.877	2	.000***
2 + focus condition	7	−982.68	1 vs 2	0.094	1	.759
3 + tone : focus condition	9	−981.80	2 vs 3	1.762	2	.414

2.4.3.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 10. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of

tone, χ^2 (2) = 24.52, p < .001. There was thus no evidence that Bai speakers varied pitch span to distinguish narrow focus from broad focus, regardless of tone.

Table 10. Pitch span, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	− 2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−869.06				
1 + tone	6	−856.80	0 vs 1	24.52	2	.000***
2 + focus condition	7	−856.80	1 vs 2	0.001	1	.974
3 + tone : focus condition	9	−854.64	2 vs 3	4.318	2	.115

2.4.3.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 11. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 32.135, p < .001. There was thus no evidence that Bai speakers varied pitch span to distinguish contrastive focus from non-contrastive, regardless of tone.

Table 11. Pitch span, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−1026.7				
1 + tone	6	−1010.6	0 vs 1	32.135	2	.00***
2 + focus condition	7	−1009.9	1 vs 2	1.559	1	.212
3 + tone : focus condition	9	−1009.6	2 vs 3	0.419	2	.811

2.4.4 Pitch maximum

2.4.4.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 12. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 58.704, p < .001. There was thus no evidence that Bai speakers varied pitch maximum to distinguish narrow focus from post-focus, regardless of tone.

Table 12. Pitch maximum, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−1162.5				
1 + tone	6	−1133.2	0 vs 1	58.704	2	.000***
2 + focus condition	7	−1133.2	1 vs 2	0.048	1	.826
3 + tone : focus condition	9	−1133.0	2 vs 3	0.231	2	.891

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 13. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 58.829, p < .001. There was thus no evidence that Bai speakers varied pitch maximum in their Bai to distinguish narrow focus from pre-focus, regardless of tone.

Table 13. Pitch maximum, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1165.0				
1 + tone	6	−1135.6	0 vs 1	58.829	2	.000***
2 + focus condition	7	−1135.4	1 vs 2	0.314	1	.575
3 + tone : focus condition	9	−1135.3	2 vs 3	0.205	2	.903

2.4.4.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 14. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, $\chi^2(2) = 53.611$, $p < .001$. There was thus no evidence that Bai speakers varied pitch maximum to distinguish narrow focus from broad focus, regardless of tone.

Table 14. Pitch maximum, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1047.3				
1 + tone	6	−1020.5	0 vs 1	53.611	2	.000***
2 + focus condition	7	−1020.5	1 vs 2	0.004	1	.953
3 + tone : focus condition	9	−1020.1	2 vs 3	0.764	2	.682

2.4.4.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 15. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 57.504, p < .001. There was thus no evidence that Bai speakers varied pitch maximum to distinguish contrastive from narrow focus (non-contrastive focus), regardless of tone.

Table 15. Pitch maximum, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	Δχ2	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1176.3				
1 + tone	6	−1147.6	0 vs 1	57.504	2	.000***
2 + focus condition	7	−1147.6	1 vs 2	0.008	1	.931
3 + tone : focus condition	9	−1147.4	2 vs 3	0.358	2	.836

2.4.5 Pitch minimum

2.4.5.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 16. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 64.702, p < .001. There was thus no evidence that Bai speakers varied pitch minimum to distinguish narrow focus from post-focus, regardless of tone.

Table 16. Pitch minimum, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	Δχ2	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1159.4				
1 + tone	6	−1127.0	0 vs 1	64.702	2	.000***
2 + focus condition	7	−1127.0	1 vs 2	0.002	1	.964
3 + tone : focus condition	9	−1126.8	2 vs 3	0.329	2	.848

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 17. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 64.233, $p < .001$. There was thus no evidence that Bai speakers varied pitch minimum to distinguish narrow focus from pre-focus, regardless of tone.

Table 17. Pitch minimum, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	Δχ2	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−1150.3				
1 + tone	6	−1118.2	0 vs 1	64.233	2	.000***
2 + focus condition	7	−1118.2	1 vs 2	0.029	1	.866
3 + tone : focus condition	9	−1117.8	2 vs 3	0.759	2	.684

2.4.5.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 18. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of

tone, χ^2 (2) = 61.667, p < .001. There was thus no evidence that Bai speakers varied pitch minimum to distinguish narrow focus from broad focus, regardless of tone.

Table 18. Pitch minimum, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−1036.9				
1 + tone	6	−1006.1	0 vs 1	61.667	2	.000***
2 + focus condition	7	−1006.1	1 vs 2	0.087	1	.769
3 + tone : focus condition	9	−1005.8	2 vs 3	0.506	2	.776

2.4.5.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 19. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (2) = 61.649, p < .001. There was thus no evidence that Bai speakers varied pitch minimum to distinguish contrastive focus from non-contrastive focus, regardless of tone.

Table 19. Pitch minimum, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−1154.1				
1 + tone	6	−1123.2	0 vs 1	61.649	2	.000***
2 + focus condition	7	−1123.2	1 vs 2	0.006	1	.939
3 + tone : focus condition	9	−1122.8	2 vs 3	0.916	2	.633

2.4.6 Interim summary

To summarize, we found that (near-) monolingual speakers of Bai only used duration to differentiate narrow focus from non-focus, regardless of tone. Specifically, the duration of the focal constituent was lengthened in narrow focus, compared to its counterpart in the post-focus condition in all lexical tones. However, duration was not used for distinguishing narrow focus from pre-focus, regardless of tone. Furthermore, we found that (near-) monolingual speakers of Bai did not vary duration to differentiate focus types that differed in constituent size and contrastiveness, regardless of tone. With regard to the use of pitch-related prosodic cues, Bai speakers did not vary any of the pitch-related cues to differentiate narrow focus from non-focus (pre-focus or post-focus) and focus types that differed in constituent size and contrastiveness, regardless of tone.

2.5 Discussion and conclusion

The present study examined prosodic focus-marking in Bai, a tone language spoken in the southwest of China, by adopting an experimental methodology. We have found that focus is prosodically encoded in Bai via duration. Specifically, (near-) monolingual speakers of Bai lengthen the duration of the focal constituent relative to its counterpart in post-focus condition to encode narrow focus, similar to speakers of Standard Mandarin (Ouyang & Kaiser, 2015; Xu, 1999; Yang, 2017; Yang & Chen, 2014). However, they do not use pitch variation in any way in focus-marking, unlike speakers of Standard Mandarin. Furthermore, speakers of this variety of Bai do not use duration to distinguish focus types that differed in the size and contrastiveness of the focal constituent, unlike speakers of Standard Mandarin (Ouyang & Kaiser, 2015; Xu, 1999; Yang & Chen, 2014), who use duration to distinguish focus types that differed in the size of the focal constituent. The results thus suggest that Bai uses prosody to mark focus, but to a lesser extent, compared to Standard Mandarin. Related to this is the fact that Bai also exploits word order and morphological topic marker to distinguish focal information from non-focal or topical information. Specifically, the canonical word order in Bai is SVO, the word order OSV can be

used to highlight the topic status of the object. Further, the topical status of a subject can be optionally marked by topic markers, such as "nɯ^{55}" and "lɯ^{44}" (Zhào, 2009). The use of these non-prosodic cues may explain the modest use of prosody in focus-marking in Bai. Although co-existence of multiple strategies does not necessarily impede the use of prosody (Chen, Lee, & Pan, 2016; Gussenhoven, 2007; Michaud & Brunelle, 2016), languages differ in the degree to which these strategies are used. For example, some languages use prosodic strategies to a larger extent than non-prosodic strategies, such as Dutch and German (Chen, 2009; Frey, 2006), but other languages use non-prosodic strategies and prosodic strategies to a similar extent, such as Finnish and Turkish (Arnhold, Chen, & Järvikivi, 2016; İşsever, 2003). Bai can thus be considered a language that uses non-prosodic strategies to a larger extent than prosodic strategies.

Together with previous studies of tone languages with different lexical tones, such as Yi, Standard Mandarin, Vietnamese, and Cantonese (Jannedy, 2007; Wang et al., 2011; Wu & Xu, 2010; Xu, 1999; Yang, 2017), we suggest that there is no relationship between the prosodic cues used to mark focus prosodically and the number of lexical tones in a language. Specifically, as we reviewed earlier, Yi and Standard Mandarin are similar in terms of having the same number of lexical tones (i.e., four), while Standard Mandarin differs from Yi in that it exploits both pitch and duration for encoding focus (Wang et al., 2011; Xu, 1999; Yang, 2017). Similarly, although both Cantonese and Vietnamese have a relatively large number of lexical tones (i.e., six), Vietnamese uses both pitch and duration for encoding focus and Cantonese only uses duration for the same purpose (Jannedy, 2007, 2008; Wu & Xu, 2010). Furthermore, our results confirmed the hypothesis that Bai is different from Standard Mandarin in the use of prosody for focus-marking purpose. This finding implies that Bai does not appear to be subjected to influence from Mandarin in prosodic focus-marking, despite the fact that Bai has been in long-lasting and intensive contact with Mandarin and has shown lexical and syntactic influence from Mandarin. Bai is thus similar to Yi, Deang, Taiwanese, and Tsat in this respect. Our study also has useful methodological implications. Previous studies have raised concerns on the representativeness of the speech elicited by asking speakers to

produce the same sentence repeatedly, mostly from a written script (Chen et al., 2014). On the other hand, it is notoriously difficult to determine information structure in semi-spontaneous speech. Our method allows us to control for information structure, to ensure consistency in choice of word, and to elicit speech that can be considered to be substantially more representative of the speech used in everyday communication than read speech. More importantly, our method does not depend on written scripts and is based on pictures. This method is therefore highly suitable for investigating speech in minority languages without a writing system and with low literacy.[22]

Finally, our study suggests two topics for future research. First, Bai has been in close contact with Mandarin for centuries (Hefright, 2011), which has led to a large number of Mandarin-loan words in Bai, and heavy influence from Mandarin syntactic structure on the syntax of Bai (Allen, 2004; Xu, 2008). However, in spite of the lexical and syntactic influence from Mandarin, Bai does not appear to be subjected to the influence from Mandarin in prosodic focus-marking. Together with existent work on Yi, Deang, and Tsat (Wang et al., 2012, 2011), this finally puts forward an interesting hypothesis to be tested in future research. That is, prosodic focus-marking may not easily undergo changes as a result of language contact. Second, considering the dialectal differences in Bai language, it is not clear whether prosodic focus-marking is similar across Bai dialects. Future research on the northern variety of Bai can shed light on this question.

[22] The interested readers are referred to www.prosodicdevelopment.com for more information on the method used in other languages and for downloading the pictures designed for the game in different languages.

CHAPTER 3 Prosodic Focus-marking in Dali Mandarin[23]

Abstract

The present study investigates prosodic marking of focus in Dali Mandarin, a variety of Xinan Guanhua (Southwestern Mandarin) spoken in Dali City, the capital of Dali Bai Autonomous Prefecture, China. Dali Mandarin as a regional variety of Mandarin has been in long-lasting and intensive contact with Bai, a Sino-Tibetan tone language. By adopting a semi-spontaneous experimental approach, subject-verb-object (hereafter SVO) Dali Mandarin sentences were elicited from (near-) monolingual speakers of Dali Mandarin in varied focus conditions. Our data showed that Dali Mandarin speakers lengthened the duration of the focal constituents, compared to their non-focal counterparts. However, they did not use duration to distinguish focus types that differed in constituent size and contrastiveness. Further, Dali Mandarin speakers did not vary pitch either for signaling focus, or for encoding focus types that differed in the size or contrastiveness of the focal constituent. These results thus suggest that Dali Mandarin speakers use only duration to mark focus, is similar to Bai but different from Standard Mandarin.

Keywords: prosody, focus, Dali Mandarin

[23] Preliminary results of a portion of the data from this chapter were presented at the 5th International Symposium on Tonal Aspects of Language, and published in the proceedings (Liu, Van de Velde, & Chen, 2016). A version of this chapter is to be submitted in article form to an appropriate journal.

3.1 Introduction

Recently, researchers have shown an increasing interest in investigating the prosodic realization of focus in dialectal/regional varieties of a language. For instance, a number of studies examined the similarities and differences of prosodic focus-marking in different regional varieties of Mandarin, including the varieties spoken in Nanchang (Gan) (Wang et al., 2011), Lan Zhou (Shen & Xu, 2016), Jinan, Liaocheng, Zibo (Duàn et al., 2013), Dalian, Harbin, Tianjin and Xi'an (Duan & Jia, 2014). In general, these investigations showed a similar manner of prosodic focus-marking in these regional varieties of Mandarin and Standard Mandarin. However, limited research on varieties of Mandarin spoken in the vicinity of a language other than Mandarin have revealed different uses of prosody in focus-marking compared to Standard Mandarin (Chen et al., 2009; Xu et al., 2012). Against this background, we examine the prosodic focus-marking in Dali Mandarin, a regional variety of Mandarin, which has been in long-lasting and intensive contact with Bai, a Sino-Tibetan tone language.

3.1.1 Focus and prosodic focus-marking

Focus has been associated with various notions (Gundel, 1999; Gussenhoven, 2008; Halliday, 1967a). In the present study, focus refers to the information in a sentence that is new to the receiver (Gundel, 1999; Lambrecht, 1994). Focus can differ in the size and contrastiveness of focal constituents. For example, the focal constituent can be on a syntactic constituent larger than a word such as a verb phrase or the whole utterance (*broad focus*) or on a single lexical word (*narrow focus*). Focus can also be distinguished in terms of the meanings it conveys, such as contrastive vs. non-contrastive focus. *Contrastive focus* refers to the situation that the focal constituent contains information that serves as a correction or forms a direct contrast to what has been previously mentioned in the discourse (Chafe, 1976).

Pitch and duration are identified as major prosodic cues for realizing focus in various languages (Chen, 2009; Cooper et al., 1985; De Jong, 1980; Xu, 1999). How pitch and duration are used exactly is, however, language-specific (Chen, 2009;

Cooper et al., 1985; Jannedy, 2007; Rump & Collier, 1996; Wang et al., 2011; Wu & Xu, 2010; Xu, 1999; Xu & Xu, 2005). For example, both duration and pitch are used for encoding focus in languages such as Standard Mandarin (Xu, 1999; Yang & Chen, 2014), Vietnamese (Jannedy, 2007), English (Cooper et al., 1985; Xu & Xu, 2005), Dutch (Chen, 2009; Rump & Collier, 1996), German (Baumann et al., 2007) and Finnish (Arnhold, 2016). But, some languages only exploit duration as a prosodic cue for marking focus, such as Cantonese (Wu & Xu, 2010), Bai (Chapter 2), and Deang (Wang et al., 2011).

3.1.2 Languages in Dali: Dali Mandarin and Bai

Dali Mandarin (*Dàlǐ fāngyán* 大理方言) is a regional variety of Southwestern Mandarin (*Xīnán guānhuà* 西南官话) spoken in Dali City (*Dàlǐshì xiàguān* 大理市下关), the capital of Dali Bai Autonomous Prefecture (*Dàlǐ báizú zìzhìzhōu* 大理白族自治州), Yunnan (*Yúnnán* 云南), China. Standard Mandarin is "widely used in the domains of science and technology, and thus has enjoyed the legal status of 'a national lingua franca' in China and is the most powerful language in the country" (Li, 2015, p. 191). The sound system of Dali Mandarin is similar to Standard Mandarin (Lǐ, 2009; Wú & Zhāng, 1988). According to the latest national survey prior to the present study (Dàlǐ báizú zìzhì zhōu dìfāngzhì biānzhuàn wěiyuánhuì bàngōngshì, 2011), the majority of the residents in the Dali area are Bai (68%) and Han (26%) people. Both Bai and Dali Mandarin are widely used, along with Standard Mandarin in Dali (Dèng & Hé, 2012; Wú & Zhāng, 1988). Bai, a Sino-Tibetan tone language, is spoken by the Bai minority ethnic group (i.e., Bai people, Báizú 白族). In the present Dali area, especially in Dali City, Dali Mandarin is mainly spoken by Han people (Hànzú 汉族). All Bai speakers are bilinguals who can speak Dali Mandarin and/or Standard Mandarin as their second language. However, there are speakers of Dali Mandarin who have little knowledge of Bai and live in the urban area of Dali City (Xiàguān 下关).

According to Wú and Zhāng's (1988) description, Dali Mandarin has nineteen phonologically distinct consonants: /pʰ, p, tʰ, t, kʰ, k, tsʰ, ts, tɕʰ, tɕ, m, n, ŋ, f, v, s, z, ɕ, x/. The velar nasal /ŋ/ appears to be used only by senior speakers of Dali Mandarin. Further, it has eight vowels:/[ɿ, i, u, y, ɐ, ʌ, o, e/,[24] four nasalized vowels /ʌ̃, ə̃, ĩ, ɛ̃/ and 22 diphthongs /ʌi, eɪ, aɔ, əu, oŋ, iʌ, io, iɛ, iu, iaɔ, iəu, iɛ̃, iʌ̃, ioŋ, uʌɪ, ue, uʌ, ueɪ, uʌ̃, yɛ, yɛ̃, y(n)/. Dali Mandarin has four lexical tones, including a high-level tone (Tone 44), a mid-falling tone (Tone 31), a high-falling tone (Tone 53) and a low-dipping tone (Tone 213). The lexical tone system of Dali Mandarin resembles the Standard Mandarin one (Lǐ, 2009; Wú & Zhāng, 1988). Specifically, Tone 44 in Dali Mandarin corresponds to Tone 1 (55) in Standard Mandarin; Tone 31 in Dali Mandarin corresponds to Tone 2 (35) in Standard Mandarin; Tone 53 in Dali Mandarin corresponds to Tone 3 (214) in Standard Mandarin; and Tone 213 in Dali Mandarin corresponds to Tone 4 (53) in Standard Mandarin.

Although the promotion of Standard Mandarin started from the 1950s in China, Bai and Mandarin have been in a long-lasting and intensive contact in the Dali area for centuries (Allen, 2004; Hefright, 2011; F. Wang, 2004; Wiersma, 2005). With the formation of an integrated China in Tang Dynasty (618–907 AD), the Bai speaking area became part of China. This close contact between Bai people and Han Chinese, and their languages—Bai and Mandarin—has continued after the foundation of the People's Republic of China in 1949. As a result of the long-lasting and intensive contact between Bai and Mandarin, Mandarin has influenced Bai so strongly that the genetic affiliation of Bai has been a subject of much debate (Hefright, 2011; F. Wang, 2004, 2005; Zhào & Xú, 1996). However, only a few observations of Bai-accented Mandarin produced by Bai-Mandarin bilinguals were made in the literature (Lǐ, 2012, pp. 67–68).

[24] In Wú and Zhāng (1988), the /ʌ/ was printed as /ʌ/. However, only the /ʌ/ symbol is listed in IPA (2015 revised version). In the following text, /ʌ/ is used in conformity with the Wú and Zhāng's (1988) description.

3.1.3 The present study

In the present study, we examine prosodic focus-marking by (near-) monolingual speakers of Dali Mandarinwho predominantly use Dali Mandarin in their daily lives,[25] and identified themselves as native speakers of Dali Mandarin. Before we probe into the details of the present study, a background in prosodic focus-marking in Bai and Standard Mandarin is provided as follows.

Bai, as a minority language widely used in Dali, only exploits duration for encoding focus (Chapter 2). It has been found that the duration of the focal constituent is lengthened to distinguish narrow focus from post-focus in Bai. However, the duration of the focal constituent is neither lengthened nor shortened relative to its counterpart in the pre-focus condition. In addition, pitch-related prosodic cues, including pitch span, pitch maximum, and pitch minimum are not varied for encoding narrow focus or focus types that differed in constituent size and contrastiveness in Bai (Chapter 2).

Unlike Bai, focus is prosodically encoded via pitch and duration in Standard Mandarin (Chen, 2010; Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015; Xu, 1999; Yang, 2017; Yang & Chen, 2014). Regarding the use of prosodic cues for encoding narrow focus, the pitch span and duration of the focal constituent are expanded and lengthened relative to the non-focal counterpart (Yang, 2017; Yang & Chen, 2017). Further, the pitch span of the post-focal constituent is compressed relative to the non-focal counterpart in broad focus condition (Xu, 1999). Regarding the use of prosodic cues for encoding focus types that differed in the size of the focal constituent, the pitch span and duration of the focal constituent are expanded and lengthened in narrow focus, relative to its counterpart in broad focus in read speech

[25] Strictly speaking, there are no monolingual speakers of Dali Mandarin, as almost all of them can speak Standard Mandarin in the context of China. We use the term "(near-) monolingual" speakers of Dali Mandarin to indicate speakers who met all the following criteria: (1) acquire Dali Mandarin as the first language from birth; (2) using Dali Mandarin on a daily basis with self-estimated daily use exceeding 60%; (3) not having lived outside the Dali Mandarin-speaking community over the last ten years at the time of testing; and (4) not having actively used other languages for a long period on a daily basis.

(Xu, 1999). However, in semi-spontaneous speech, only the duration of the focal constituent is lengthened in narrow focus to distinguish from broad focus, while the pitch span is not varied (Yang, 2017; Yang & Chen, 2014). Regarding the use of prosodic cues for encoding focus types that differed in the contrastiveness of the focal constituent, the pitch span and duration of the focal constituent are expanded and lengthened in contrastive focus, relative to its counterpart in non-focus in read speech (Chen & Gussenhoven, 2008). However, in semi-spontaneous speech, neither pitch span nor duration is used to distinguish contrastive focus from narrow focus in semi-spontaneous speech (Yang, 2017; Yang & Chen, 2017).

It has been observed that many regional varieties of Mandarin make use of pitch and duration for encoding focus in a similar manner as in Standard Mandarin, such as regional varieties of Mandarin spoken in Nanchang (Gan dialect) (Wang et al., 2011), Lanzhou (Shen & Xu, 2016), Jinan, Liaocheng, Zibo (Duàn et al., 2013), Dalian, Harbin, Tianjin and Xi'an (Duan & Jia, 2014). Namely, speakers of the abovementioned varieties of Mandarin encode focus either by raising the mean pitch of the focal constituent or by expanding the pitch span relative to the non-focal counterpart. In addition, the compression of the pitch span of the post-focal constituent has also been consistently found in all abovementioned varieties of Mandarin. However, some differences in the use of duration between these varieties of Mandarin are found. For example, in Lan-Yin Mandarin (a regional variety of Mandarin spoken in Lanzhou), on-focus duration change was only present in sentence-medial words (Shen & Xu, 2016). Duàn et al (2013) found that in regional varieties of Mandarin spoken in Jinan, Liaocheng, and Zibo, no noticeable difference in syllable duration between the focused and unfocused conditions was found. Further, Duan and Jia (2014) also did not find the systematic variations in syllable duration of focused constituents in comparison to unfocused constituents in Dalian, Harbin, Jinan, Tianjin and Xi'an Mandarin. Despite the differences in the use of duration, it is quite clear that pitch plays a major role in signalling focus in various regional varieties of Mandarin and Standard Mandarin.

However, limited research on varieties of Mandarin spoken in the vicinity of a language other than Mandarin have revealed different uses of prosody in focus-marking compared to Standard Mandarin. For example, Xu et al. (2012) investigated focus-marking in Taiwan Mandarin, compared to Standard Mandarin and Taiwanese. Taiwan Mandarin, spoken in Taiwan, has been in close contact with Taiwanese for several decades (Chen et al., 2009). It has been shown that Taiwan Mandarin is more similar to Taiwanese than to Standard Mandarin in terms of using prosodic cues for marking focus. Specifically, Taiwan Mandarin monolinguals not only expanded pitch span but also lengthened the duration of the focal constituents to encode focus as Standard Mandarin and Taiwanese speakers did. However, neither Taiwan Mandarin speakers nor Taiwanese speakers produced the post-focal constituents with compressed pitch span, which was present in the Standard Mandarin speakers' production. It was suggested that these differences between Taiwan Mandarin and Standard Mandarin could be contributed to Taiwan Mandarin's close contact with Taiwanese.

Considering that varieties of Mandarin spoken in the vicinity of a language other than Mandarin showed different uses of prosody for marking focus compared to Standard Mandarin, such as Taiwan Mandarin (Xu et al., 2012), and that Dali Mandarin has been in long-lasting and intensive contact with Bai, we hypothesize that Dali Mandarin exploits prosody for encoding focus in a similar manner as in Bai. The predictions are that (near-) monolingual speakers of Dali Mandarin will lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and that they will vary neither pitch-related cues to encode narrow focus nor duration to encode focus types that differed in constituent size, like speakers of Bai.

3.2 Method

3.2.1 Picture-matching game

A picture-matching game paradigm used Yang and Chen (Yang & Chen, 2017, 2014) and Yang (2017) was adopted to ensure the comparability between the present data and the only data on semi-spontaneous speech in Standard Mandarin (Yang & Chen, 2017). The participants of the present study were informed that the picture-matching game was designed so that it could be used in research on children's prosodic focus-marking, and consequently, the task could be quite simple for adults.

In the picture-matching game, three sets of pictures were used. The experimenter and the participant each held a set of pictures ordered in a specific sequence; and the third set of pictures was scattered around on a table. In the experimenter's pictures (the first set), there was always some information missing, e.g., a subject, an action (verb), an object, or all three. The participant's pictures (the second set) all contained a complete event. The participant's task was to help the experimenter with sorting out pictures from the first set (experimenter's pictures) and the third set (scattered around on the table) in matched pairs. An example of a trial eliciting a target sentence with **n**arrow-**f**ocus on a verb in sentence-**m**edial position (the NF-m condition): First, the experimenter took a picture (e.g., a dog and a ball) from her own set, drew the participant's attention to the picture and established what the picture was by saying, e.g., "Look! The dog and the ball. It looks like the dog does something to the ball." This was done to make sure that the entity in the picture was referentially given to the participant before the utterance of the question. Second, the experimenter asked a question about the picture (e.g., "What does the dog do to the ball?"). Third, the participant took a picture from his or her set and inspected it. The experimenter then repeated the question, followed by an answer from the participant (e.g., "The dog [THROWS] the ball."). Lastly, the experimenter found the picture containing the missing information in the third set and paired it up with her own picture. As to rules of the game, participants were explicitly asked to respond in full sentences and not to show their pictures to the experimenter. To ensure consistency in the participants' word choice, the picture-matching game was preceded by a

picture-naming task, which was designed to familiarize the participants with the target words and entities in the pictures used in the game.

3.2.2 Experimental material

SVO sentences in five focus conditions were elicited via the picture-matching game: **n**arrow-**f**ocus on the subject NP in sentence-**i**nitial position (NF-i); **n**arrow-**f**ocus on the verb in sentence-**m**edial position (NF-m); **n**arrow-**f**ocus on the object NP in sentence-**f**inal position (NF-f); **b**road **f**ocus (BF) and **c**ontrastive-**f**ocus on the verb in sentence-**m**edial position (CF-m). The focus condition was set up by a WH-question or a statement from the experimenter, as illustrated in examples (1) to (5).

(1) Experimenter: Look! This picture is very blurry. I can't see anything clearly. What does the picture show?
Participant: [THE BEAR THROWS THE BALL.]
(BF: BROAD FOCUS)

(2) Experimenter: Look! The ball. There is also a waving arm. It looks like someone throws the ball. Who throws the ball?
Participant: [THE BEAR] throws the ball.
(NF-I: NARROW FOCUS ON THE SUBJECT)

(3) Experimenter: Look! The dog and the ball. It looks like the dog does something with the ball. What does the dog do to the ball?
Participant: The dog [THROWS] the ball.
(NF-M: NARROW FOCUS ON THE VERB)

(4) Experimenter: Look! The rabbit, it waves its arm. It looks like the rabbit throws something. What does the rabbit throw?
Participant: The rabbit throws [THE BALL].
(NF-F: NARROW FOCUS ON THE OBJECT)

(5) Experimenter: Look! The cat and the ball. It looks like the cat does something to the ball. I guess the cat cuts the ball.

Participant: The cat [THROWS] the ball.

(CF-M: CONTRASTIVE FOCUS ON THE VERB)

The SVO sentences were constructed in such a way that each was a unique combination of a subject-noun and a VP (verb + object-noun). Each of the four Dali Mandairn lexical tones (i.e., high-level tone/Tone 44, mid-falling tone/Tone 31, high-falling tone/Tone 53, and low-dipping tone/Tone 213) occurred in subject noun phrases, verbs, and object nouns. Four monosyllabic verbs and four monosyllabic object nouns formed 16 VPs, each of which appeared in each focus condition (n = 5). This resulted in 80 VPs, and then four subject-noun phrases, which started with the word "xiǎo (little)" and ended with four nouns carrying four lexical tones were evenly distributed over these 80 VPs to form 80 SVO target sentences. An overview of the words in SVO sentences is provided in Table 1.

Table 1. Overview of the words in SVO sentences, each word is presented in Chinese, IPA and English

	Subject	Verb	Object
Tone 44	小猫 ɕiao^{53}maɔ^{44} cat (little cat)	扔 zɜ̃^{44} throw	书 su^{44} book
Tone 31	小熊 ɕiao^{53}ɕioŋ^{31} bear (little bear)	埋 mʌi^{31} bury	球 tɕ^{h}iəu^{31} ball
Tone 53	小狗 ɕiao^{53}kəu^{53} dog (little dog)	剪 tɕiɜ̃^{53} cut	笔 pi^{53} pen
Tone 213	小兔 ɕiao^{53}t^{h}u^{213} rabbit (little rabbit)	运 yn^{213} transport	菜 ts^{h}ʌi^{213} vegetable

3.2.3 Participants and procedure

Six (near-) monolingual speakers of Dali Mandarin (three male and three female, age range: 28 to 54 years, mean age = 34 years, SD = 10.1) participated in our study. Participants predominantly used Dali Mandarin in their daily lives, identified themselves as native speakers of Dali Mandarin, and acquired Standard Mandarin as their second language at the age of six (see Appendix D). The participants all met the following criteria: (1) using Dali Mandarin on a daily basis with self-estimated daily use exceeding 60%; (2) not having lived outside the Dali Mandarin-speaking community over the last 10 years at the time of testing; (3) not having actively used Bai or other languages for a long period on a daily basis; and (4) having no self-reported speech and hearing impairments.

The participants were tested individually by a female experimenter, who was a native speaker of Dali Mandarin (age = 27) and only spoke Dali Mandarin with the participants. Each experiment was split into two sessions, each contained 40 trials. A break was introduced after one session to avoid loss of attention or fatigue in the participants. Each session took 25 to 30 minutes for a participant to complete. All the test sessions were conducted either in a quiet room at the experimenter's apartment or the participant's home in Dali City, Dali Bai Autonomous Prefecture, China. The experiments were recorded using a portable ZOOM H1 digital recorder at a 44.1 kHz sampling rate and 16-bit accuracy. Each session was also video-recorded for future training purpose.

3.2.4 Acoustic annotation

The auditory recordings from the participants were first orthographically annotated. An answer sentence was discarded for analysis if had one of the following features: (1) it was not uttered as a direct response to the target question; (2) it contained self-corrections; (3) it contained hesitations (defined as a long "em" sound produced by the participants before answering questions); (4) it deviated from the target sentence in word choice or word order. In total, 92% of the obtained responses (n = 440) were selected for the analysis.

The **verbs** were the targets of prosodic analysis. They were focal constituents in the NF-m, BF and CF-m conditions, post-focal constituents in the NF-i condition, and pre-focal constituents in the NF-f condition. The verbs were acoustically annotated by examining the waveform, wide-band spectrum, and pitch track in Praat (Boersma & Weenink, 2006) in combination with auditory impressions (Turk et al., 2006). Two pitch-related landmarks and two segmental landmarks were labelled in each verb: pitch maximum, pitch minimum, word onset, and word offset. An example of the annotation is presented in Figure 1.

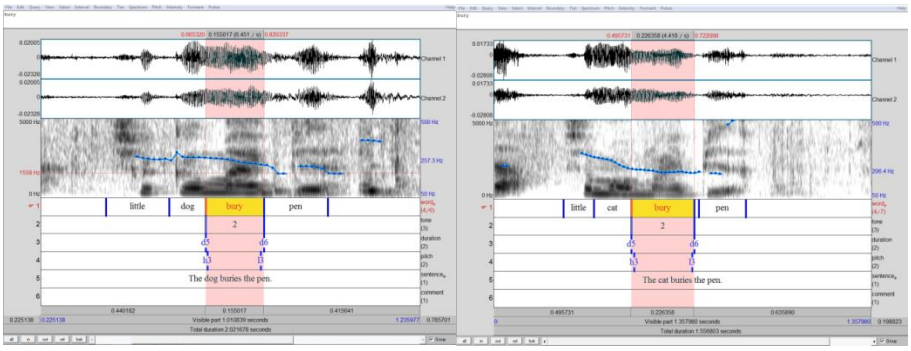

Figure 1. Pitch contour (in Hz) of verbs in narrow focus vs. contrastive focus position (NF-m = verb in sentence medial focused position, CF-m = verb in sentence medial contrastive focused constituent).

The tonal targets of the lexical tones in Dali Mandarin were taken into consideration when the landmarks for pitch maximum were inserted (Xu, 1997; Xu & Wang, 2001; Yang, 2017; Yang & Chen, 2014). Specifically, Tone 44 is a high-level tone and often appeared as a slightly rising pitch contour and occasionally as a slightly falling pitch contour sentence-medially, as reported for its counterpart (Tone 1) in Standard Mandarin in Xu (1997) and Yang and Chen (2014). The pitch-maximum was labelled after the pitch-minimum in the case of a rising Tone 44 syllable but before the pitch-minimum in the case of a falling Tone 44 syllable. Tone 31 is a falling tone, but we observed that it appeared as a rise-fall contour sentence-medially, again, as reported for its counterpart (Tone 4) in Standard Mandarin

described in Xu (1997) and Yang and Chen (2014). As the rising part is subjected to the influence of the preceding tone, the falling part contains the tonal targets of Tone 31. Thus, the pitch-maximum was labelled before the pitch-minimum in the falling part of Tone 31. Tone 53 is also a falling tone, but we observed that Tone 53, a falling tone in its citation form, could be realized as either a rise (66%, n = 72) or a fall (34%, n = 37) due to the tonal context (n_{total} = 109). In the case of a rising Tone 53, we labelled the pitch maximum landmark after the pitch minimum landmark; in the case of a falling Tone 53, we labelled the pitch maximum before the pitch minimum landmark. For Tone 213, we observed that it was always realized as a low falling tone and could be transcribed as Tone 21 rather than Tone 213. Thus, the pitch maximum landmark was labelled before the pitch minimum landmark in Tone 213. In the following text, Tone 213 is used in conformity with the previous description.

The pitch values (in Hz) at the pitch-related landmarks, and time values (in seconds) at the segmental landmarks were automatically extracted by means of Praat scripts. Four measures were obtained from each verb: pitch maximum, pitch minimum, pitch span (i.e., the difference between the maximum pitch and the minimum pitch), and word duration. In 25 of the usable responses (6%), an accurate measurement of pitch values was not possible. Thus, these sentences were excluded from the analysis of pitch-related measurements.

3.3 Statistical analyses and results

3.3.1 Statistical analyses

In order to assess the effect of focus, we compared the measurements of the verbs between the unfocused conditions and the focused condition; that is, NF-m (verb is the focal constituent) vs. NF-i (verb is the post-focal constituent), and NF-m (verb is the focal constituent) vs. NF-f (verb is the pre-focal constituent). To investigate the effect of focus type that differed in the size of the focal constituent, we compared narrow focus on the verb (NF-m) with broad focus (BF). To investigate the effect of

focus type that differed in the contrastiveness of the focal constituent, we compared contrastive focus on the verb (CF-m) with non-contrastive focus on the verb (NF-m).

Statistical analyses were conducted using mixed-effects modelling in R (R Core Team, 2014) with package "lme4" (Bates, Mächler, et al., 2015) and "lmerTest" (Kuznetsova et al. 2013). In all models, tone and focus conditions were included as fixed factors, while speaker (i.e., the participants) and sentence (i.e., the answer sentences) were included as random factors. In each of the four pair-wise comparisons listed above, focus condition had two levels (the two focus conditions of interest), and tone referred to the four lexical tones of the target verbs (Tones 44, 31, 53, and 213). Outcome variables were the duration, pitch span, pitch maximum, and pitch minimum of the target verbs.

Following Magezi (2015) and Field et al. (2012), our models were constructed and evaluated in a stepwise fashion. Specifically, we started from an intercept-only model, built the subsequent models by adding one new term/factor at a time, and then systematically compared the models which differed only by the presence or absence of one term. The comparison was done using a likelihood-ratio test, and the test statistic χ^2, degrees of freedom and *p*-value were reported for the added term. A *p*-value of less than 0.05 was considered to indicate that the added term contributed significantly to the model fit (Magezi, 2015). A detailed description of the model building is presented in Table 2.

Table 2. Model build-up procedure

Model	Added term
Model 0	only contains speaker, sentence as random intercepts
Model 1	+ tone
Model 2	+ focus condition
Model 3	+ tone : focus condition

When building the models, only the factors and interactions that significantly improved the fit of the model were retained until the best-fit model was determined. After the best-fit model was established, we only summarized and interpreted the model with the best fit. For each analysis, we reported the results of the model comparison first, and then the parameter estimates of the best-fit model. As we are primarily interested in the effect of focus on the outcome variables, we will concentrate on the main effect of focus or interactions involving the factor of focus. However, we do not discuss in detail the main effects of the factors when the interactions involving these factors are significant.

3.3.2 Duration

3.3.2.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The effect of focus (narrow focus vs. post-focus) on duration was tested with the models specified in Table 2. The results of the fit and model comparisons are shown in Table 3. As can be seen, the best-fit model was the model only containing the main effect of focus condition, χ^2 (1) = 5.376, $p < .05$. Parameter estimates of the best-fit model are presented in Table 4. The main effect of focus condition was such that the duration of the verbs was significantly longer in narrow focus (226.1 ms, SD = 44.8) than in post-focus (213.2 ms, SD = 30.5), regardless of tone (b = 14.653, df = 30.362, t = 2.405, $p < .05$), as shown in Figure 2.

Table 3. Duration, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−860.83				
1 + tone	7	−860.62	0 vs 1	0.418	3	.937
2 + focus condition	8	−857.93	1 vs 2	5.376	1	.020*
3 + tone : focus condition	11	−857.82	2 vs 3	0.223	3	.974

Table 4. Duration, narrow focus (NF-m) vs. post-focus (NF-i), parameter estimates of the best-fit model

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	213.109	11.095	7.673	19.207	.000***
Narrow focus	14.653	6.092	30.362	2.405	.023*
Random part	*Name*	S^2	*SE*		
Sentence	Intercept	179.7	13.41		
Speaker	Intercept	628.6	25.07		
Residual		648.8	25.47		

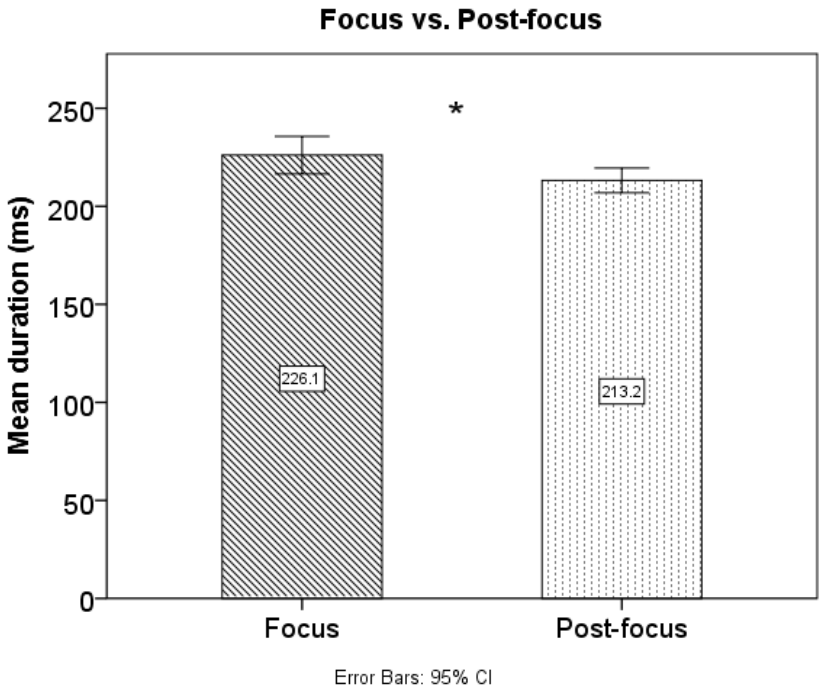

Figure 2. Mean duration (in ms) of focal constituents vs. post-focal constituents, n = 179, N = 6. Significant differences are marked with an asterisk.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 5. As can be seen, the best-fit model was the model only containing the main effect of focus condition, χ^2 (1) = 8.139, $p < .01$. Parameter estimates of the best-fit model are presented in Table 6. The main effect of focus condition was such that the duration of the verbs was significantly longer in narrow focus (226.1 ms, SD = 44.8) than in pre-focus (208.3 ms, SD = 36.6), regardless of tone (b = 19.608, df = 30.503, t = 3.05, $p < .01$), as shown in Figure 3.

Table 5. Duration, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−849.88				
1 + tone	7	−849.74	0 vs 1	0.282	3	.963
2 + focus condition	8	−845.67	1 vs 2	8.139	1	.004**
3 + tone : focus condition	11	−845.07	2 vs 3	1.209	3	.751

Table 6. Duration, narrow focus (NF-m) vs. pre-focus (NF-f), parameter estimates of the best-fit model

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	208.575	12.853	7.453	16.23	.000***
Narrow focus	19.608	6.429	30.503	3.05	.005**
Random part	*Name*	S^2	*SE*		
Sentence	Intercept	228.6	15.12		
Speaker	Intercept	868.2	29.46		
Residual		561.6	23.70		

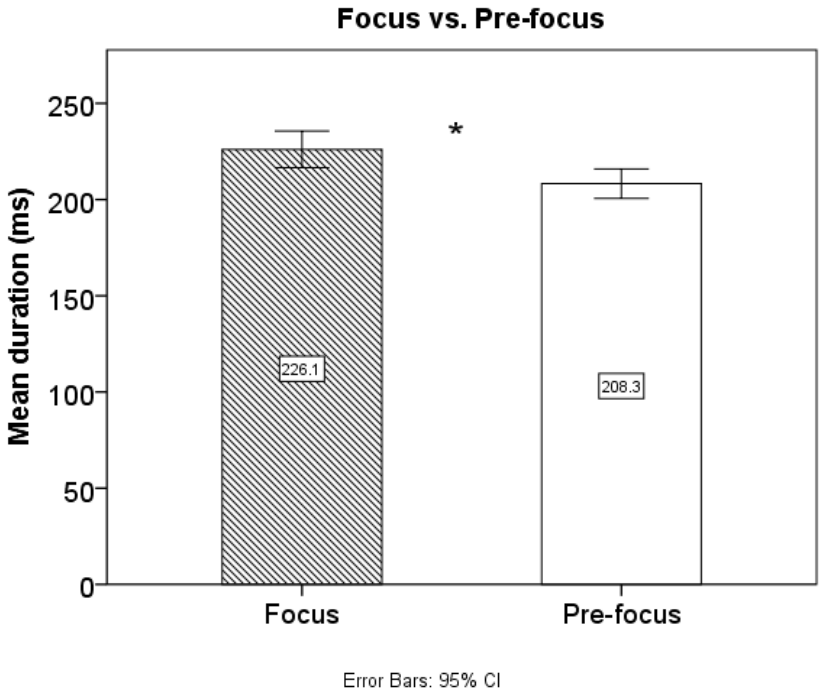

Figure 3. Mean duration (in ms) of focal constituents vs. pre-focal constituents, n = 178, N = 6. Significant differences are marked with an asterisk.

3.3.2.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The effect of size (narrow focus vs. broad focus) on duration was tested with the models specified in Table 2. The results of the fit and model comparisons are shown in Table 7. As can be seen, none of the fixed factors could improve the fit of the model. There was thus no evidence that Dali Mandarin speakers varied duration to distinguish narrow focus from broad focus, regardless of tone.

Table 7. Duration, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−827.48				
1 + tone	7	−827.35	0 vs 1	0.255	3	.968
2 + focus condition	8	−825.48	1 vs 2	3.736	1	.053.
3 + tone : focus condition	11	−825.33	2 vs 3	0.289	3	.962

3.3.2.3 Effect of contrastiveness: contrastive focus (CF-m) vs narrow focus (NF-m)

The effect of contrastiveness (contrastive focus vs. narrow focus) on duration was tested with models specified in Table 2. The results of the fit and model comparisons are shown in Table 8. As can be seen, none of the fixed factors could improve the fit of the model. There was thus no evidence that Dali Mandarin speakers varied duration to distinguish contrastive focus from non-contrastive focus (i.e., narrow focus), regardless of tone.

Table 8. Duration, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−824.68				
1 + tone	7	−824.68	0 vs 1	0.199	3	.978
2 + focus condition	8	−824.16	1 vs 2	0.829	1	.363
3 + tone : focus condition	11	−823.97	2 vs 3	0.391	3	.942

3.3.3 Pitch span

3.3.3.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The effect of focus (narrow focus vs. post-focus) on pitch span was tested with the models specified in Table 2. The results of the fit and model comparisons are shown in Table 9. As can be seen, the best-fit model was the model containing the main effect of tone, χ^2 (3) = 11.019, p < .05. There was thus no evidence that Dali Mandarin speakers varied pitch span to distinguish narrow focus from post-focus, regardless of tone.

Table 9. Pitch span, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−707.52				
1 + tone	7	−702.01	0 vs 1	11.019	3	.012*
2 + focus condition	8	−701.64	1 vs 2	0.750	1	.387
3 + tone : focus condition	11	−701.23	2 vs 3	0.813	3	.847

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 10. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 15.852, p < .01. There was thus no evidence that Dali Mandarin speakers varied pitch span to distinguish narrow focus from pre-focus, regardless of tone.

Table 10. Pitch span, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−702.78				
1 + tone	7	−694.86	0 vs 1	15.852	3	.001**
2 + focus condition	8	−694.71	1 vs 2	0.288	1	.592
3 + tone : focus condition	11	−694.67	2 vs 3	0.086	3	.994

3.3.3.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The effect of size (narrow focus vs. broad focus) on pitch span was tested with the models specified in Table 2. The results of the fit and model comparisons are shown in Table 11. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 13.873, $p < .01$. There was thus no evidence that Dali Mandarin speakers varied pitch span to distinguish narrow focus from broad focus, regardless of tone.

Table 11. Pitch span, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−675.42				
1 + tone	7	−668.48	0 vs 1	13.873	3	.003**
2 + focus condition	8	−668.47	1 vs 2	0.025	1	.875
3 + tone : focus condition	11	−668.14	2 vs 3	0.656	3	.885

3.3.3.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The effect of contrastiveness (contrastive focus vs. narrow focus) on pitch span was tested with the models specified in Table 2. The results of the fit and model comparisons are shown in Table 12. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 17.887, p < .001. There was thus no evidence that Dali Mandarin speakers varied pitch span to distinguish contrastive focus from non-contrastive focus (i.e., narrow focus), regardless of tone.

Table 12. Pitch span, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−680.40				
1 + tone	7	−671.46	0 vs 1	17.887	3	.000***
2 + focus condition	8	−671.45	1 vs 2	0.013	1	.911
3 + tone : focus condition	11	−671.40	2 vs 3	0.118	3	.990

3.3.4 Pitch maximum

3.3.4.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 13. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 18.773, p < .001. There was thus no evidence that Dali Mandarin speakers varied pitch maximum to distinguish narrow focus from post-focus, regardless of tone.

Table 13. Pitch maximum, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−759.21				
1 + tone	7	−749.83	0 vs 1	18.773	3	.000***
2 + focus condition	8	−749.55	1 vs 2	0.541	1	.462
3 + tone : focus condition	11	−749.28	2 vs 3	0.543	3	.909

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 14. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 20.811, $p < .001$. There was thus no evidence that Dali Mandarin speakers varied pitch maximum to distinguish narrow focus from pre-focus, regardless of tone.

Table 14. Pitch maximum, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−749.57				
1 + tone	7	−739.16	0 vs 1	20.811	3	.000***
2 + focus condition	8	−739.15	1 vs 2	0.027	1	.869
3 + tone : focus condition	11	−738.84	2 vs 3	0.621	3	.892

3.3.4.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 15. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 20.282, p < .001. There was thus no evidence that Dali Mandarin speakers varied pitch maximum to distinguish narrow focus from broad focus, regardless of tone.

Table 15. Pitch maximum, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−722.51				
1 + tone	7	−712.37	0 vs 1	20.282	3	.000***
2 + focus condition	8	−712.37	1 vs 2	0	1	.995
3 + tone : focus condition	11	−712.07	2 vs 3	0.587	3	.899

3.3.4.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 16. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 27.072, p < .001. There was thus no evidence that Dali Mandarin speakers varied pitch maximum to distinguish contrastive focus from non-contrastive focus, regardless of tone.

Table 16. Pitch maximum, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−725.66				
1 + tone	7	−712.12	0 vs 1	27.072	3	.000***
2 + focus condition	8	−712.04	1 vs 2	0.161	1	.688
3 + tone : focus condition	11	−711.93	2 vs 3	0.220	3	.974

3.3.5 Pitch minimum

3.3.5.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 17. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 53.179, $p < .001$. There was thus no evidence that Dali Mandarin speakers varied pitch minimum to distinguish narrow focus from post-focus, regardless of tone.

Table 17. Pitch minimum, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−747.91				
1 + tone	7	−721.33	0 vs 1	53.179	3	.000***
2 + focus condition	8	−721.32	1 vs 2	0.000	1	.976
3 + tone : focus condition	11	−719.38	2 vs 3	3.886	3	.274

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 18. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 62.169, p < .001. There was thus no evidence that Dali Mandarin speakers varied pitch minimum to distinguish narrow focus from pre-focus, regardless of tone.

Table 18. Pitch minimum, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−739.90				
1 + tone	7	−708.82	0 vs 1	62.169	3	.000***
2 + focus condition	8	−708.26	1 vs 2	1.121	1	.290
3 + tone : focus condition	11	−706.43	2 vs 3	3.651	3	.302

3.3.5.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 19. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 57.995, p < .001. There was thus no evidence that Dali Mandarin speakers varied pitch minimum to distinguish narrow focus from broad focus, regardless of tone.

Table 19. Pitch minimum, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−722.43				
1 + tone	7	−693.43	0 vs 1	57.995	3	.000***
2 + focus condition	8	−693.30	1 vs 2	0.260	1	.610
3 + tone : focus condition	11	−692.74	2 vs 3	1.128	3	.770

3.3.5.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 20. As can be seen, the best-fit model obtained was Model 1, which contained only the main effect of tone, χ^2 (3) = 59.987, $p < .001$. There was thus no evidence that Dali Mandarin speakers varied pitch minimum to distinguish contrastive focus from non-contrastive focus (i.e., narrow focus), regardless of tone.

Table 20. Pitch minimum, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−716.19				
1 + tone	7	−686.20	0 vs 1	59.987	3	.000***
2 + focus condition	8	−686.08	1 vs 2	0.245	1	.621
3 + tone : focus condition	11	−685.72	2 vs 3	0.711	3	.871

3.4 Discussion and conclusion

We have examined the prosodic realization of focus and focus types in Dali Mandarin semi-spontaneous speech, taking the lexical tone system into consideration. Our results show that focus is prosodically realized via duration in Dali Mandarin, similar to Bai but different from Standard Mandarin, as illustrated in an overview of the prosodic realization of focus in Bai (Chapter 2), Dali Mandarin (present chapter), and Standard Mandarin (Yang, 2017; Yang & Chen, 2017) in Table 21. These three studies employed the same data elicitation method and are thus ideal for comparison.

Table 21. Overview of the prosodic realization of focus in Bai (Chapter 2), Dali Mandarin (present chapter) and Standard Mandarin (Yang, 2017; Yang & Chen, 2017)

Effect	Comparison	Language varieties	Duration	Pitch span	Pitch-max	Pitch-min
Focus	Focus vs. Pre-focus	Standard Mandarin	+	+	-	+
		Dali Mandarin	+	-	-	-
		Bai	-	-	-	-
	Focus vs. Post-focus	Standard Mandarin	+	+	+	> (T1)
		Dali Mandarin	+	-	-	-
		Bai	+	-	-	-
Size	Narrow vs. Broad	Standard Mandarin	+	-	-	-
		Dali Mandarin	-	-	-	-
		Bai	-	-	-	-
Contrastiveness	Narrow vs. Contrastive	Standard Mandarin	-	-	-	-
		Dali Mandarin	-	-	-	-
		Bai	-	-	-	-

Regarding the prosodic realization of narrow focus in Dali Mandarin, the duration of the focal constituent is lengthened relative to its counterpart in the post-focus and pre-focus conditions, but pitch-related cues, including pitch span, pitch maximum, and pitch minimum, are not systematically used. Dali Mandarin thus differs from Standard Mandarin but resembles Bai in the manner of prosodically realizing narrow focus. Furthermore, neither duration nor pitch-related prosodic cues are used for differentiating focus types that differed in constituent size in Dali Mandarin, which is similar to Bai (Chapter 2) but different from Standard Mandarin (Yang, 2017; Yang & Chen, 2017). Finally, neither duration nor pitch-related prosodic cues are used for differentiating focus types that differed in the contrastiveness of the focal constituent in Dali Mandarin, similar to Bai (Chapter 2) and Standard Mandarin (Yang, 2017; Yang & Chen, 2017).

Therefore, our results confirm our hypothesis that Dali Mandarin, which has been in long-lasting and intensive contact with Bai, exploits prosody for encoding focus in a similar manner as Bai. Unlike many regional varieties of Mandarin and Standard Mandarin (Chen, 2010; Duan & Jia, 2014; Duàn et al., 2013; Ouyang & Kaiser, 2015; Shen & Xu, 2016; Wang et al., 2011; Xu, 1999; Yang, 2017), focus is prosodically encoded via duration in Dali Mandarin. Together with findings on Taiwan Mandarin (Chen et al., 2009; Xu et al., 2012), our finding shows that close contact with a less prestigious language can lead to remarkable differences in prosodic focus-marking between regional varieties of Mandarin and Standard Mandarin.

CHAPTER 4 Prosodic Focus-marking in Minority L1 Bai-Children Learning Mandarin as L2[26]

Abstract

This study examines the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children's Mandarin. Subject-verb-object (hereafter SVO) sentences were elicited from Bai-Mandarin bilinguals aged six to thirteen in Mandarin in varied focus conditions in a semi-spontaneous setting. The use of prosodic cues including duration, pitch span, pitch maximum, and pitch minimum for encoding focus was examined in different lexical tones. We have found that Bai-Mandarin early bilingual children mastered duration as a prosodic cue for encoding narrow focus at the age of six to seven years. However, they did not acquire the use of pitch for encoding narrow focus, even at the age of twelve to thirteen years (i.e., after five years of formal Standard Mandarin education). Additionally, there was no evidence that Bai-Mandarin early bilingual children used prosody for encoding focus types that differed in the size and contrastiveness of the focal constituent in Mandarin. These results thus show similarities and differences in the acquisition rate and route in the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children's L2 and that in monolingual Mandarin-speaking peers' L1.

Keywords: early bilingualism, prosody, focus, Bai, Mandarin Chinese

[26] Preliminary results of a portion of the data from this chapter were presented at the 40th Annual Boston University Conference on Language Development, and published in the proceedings (Liu, Chen, & Van de Velde, 2016b).

4.1 Introduction

There is a large population of children who grow up with more than one language, both globally (Grosjean, 2010; Tucker, 1998) and domestically in China.[27] However, the current understanding of language development in bilingual children is still limited, especially in the domain of prosody (Hoffmann, 2014; Paradis, 2007). This study concerns the acquisition of a cross-linguistically common function of prosody, that is, to highlight new information or to mark focus in a sentence (Gundel, 1999; Gussenhoven, 2007), in early bilinguals' L2. Following the age criteria set in Unsworth (2005), we define early bilinguals as learners whose acquisition of the second language (hereafter L2) begins between four to seven years, and late bilinguals as learners whose acquisition of the second language begins at the age of eight years or older.[28]

A central question in research on early bilingual children's L2 acquisition is whether it resembles monolingual children's L1 acquisition (Meisel, 2004; Paradis, 2007; Unsworth, 2005). The answers seem to depend on the modality (e.g., production vs. comprehension) and specific linguistic domain under investigation, such as lexicon, syntax, and segmental phonology (Dulay & Burt, 1974; Jia, 2003; Meisel, 2008; Paradis, 2007; Tsukada et al., 2004; Unsworth, 2005). The question of whether early bilingual children's L2 acquisition resembles monolingual children's L1 acquisition in terms of the developmental trajectory of prosodic focus-marking remains to be investigated. For example, previous studies suggested that early bilingual children's L2 phonological acquisition (e.g., acquisition of word-final voiceless stops) is not comparable to age-matched monolingual children'L1 phonological acquisition (Baker & Trofimovich, 2005; Trofimovich & Baker, 2007; Tsukada et al., 2004). However, many studies have provided evidence that the route and rate of early

[27] Mandarin is spoken by the majority of the Chinese population. However, there are 55 minority ethnic groups (one billion people) in China, and 52 of the minority ethnic groups have their own languages.

[28] "Early bilinguals" defined in the present study is an equivalent term to "child L2 learners" defined in Paradis (2007). Specifically, child L2 learners are denoted as children "who have established one language before they begin learning the other, and typically speak the L1 at home and the L2 at school" (Paradis, 2007, p. 387).

bilinguals' L2 acquisition are similar to those of monolingual children's L1 (Dulay & Burt, 1973, 1974; Jia, 2003; Krashen, 1982; Paradis, 2005, 2007). For example, Paradis (2005) examined the English vocabulary knowledge of early bilingual children who had various minority languages as L1s and acquired English as an L2 in Canada and found that the early bilingual children gained twelve months developmentally in vocabulary knowledge within twelve chronological months of exposure to their L2. Thus, the overall rate of vocabulary accumulation for early bilingual children's L2 acquisition is similar to the rate for L1 acquisition. Further, Jia (2003) found that the average time of exposure for Mandarin (L1)-English (L2) early bilingual children to master the use of plural "-s" in their L2-English was similar to the average time of exposure for English-speaking monolingual children's L1 acquisition. Regarding the route of acquisition, Dulay and Burt (1973, 1974) and Krashen (1982) found that the acquisition of morphemes such as progressive "-ing" and plural "-s" were acquired earlier than past tense "-ed" and third person singular "-s" in early bilingual children's L2-English development, as in monolingual English-speaking children's L1 development.

Recent years have seen increasing interest in the developmental trajectory of prosodic focus-marking in monolingual children (Arnhold, 2016; Chen, 2009, 2011; Grigos & Patel, 2010; Müller, Höhle, Schmitz, & Weissenborn, 2006; Romøren, 2016; Wonnacott & Watson, 2008; Yang & Chen, 2014). However, research on prosodic focus-marking in a bilingual context primarily focuses on adults' competence, thus not the developmental aspect (Barnes & Michnowicz, 2015; Bullock, 2009; Colantoni, 2011; Colantoni & Gurlekian, 2004; Grosser, 1997; Gut & Pillai, 2014; O'Rourke, 2005, 2012; Swerts & Zerbian, 2010; Van Rijswijk & Muntendam, 2012; Van Rijswijk et al., 2017; Zerbian, 2013). Questions such as how early bilingual children acquire prosodic focus-marking in their L2, and whether early bilingual children's L2 acquisition show the same rate and route of acquisition as monolingual children's L1 acquisition remain to be investigated. Against this background, the present study focuses on the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children's Mandarin, with the aim to

elucidate the similarities and differences between early bilingual children's L2 acquisition and monolingual children's L1 acquisition in the domain of prosody.

In the rest of the section, we will fist briefly consider the notion of focus, cross-linguistic similarities and differences in prosodic focus-marking, and how these can shape the acquisition of prosodic focus-marking in children acquiring different languages (section 4.1.1). Then we will discuss prosodic focus-marking in monolingual Mandarin-speaking adults and children in more detail in section 4.1.2. Finally, we will present the research question, hypotheses and predictions in section 4.1.3.

4.1.1 Focus and acquisition of prosodic focus-marking across languages

Focus has been associated with various notions, such as "new" and "emphasis" (Gundel, 1999; Gussenhoven, 2008; Halliday, 1967a). In the current study, we use "focus" to refer to the new information to the receiver in a sentence (Gundel, 1999; Lambrecht, 1994). Focus can differ in the size and contrastiveness of focal constituents. For example, the focal constituent can be on a single lexical word (*narrow focus*) or on a syntactic constituent larger than a word, such as a verb phrase or the whole utterance (*broad focus*). The focal constituent can contain information that serves as a correction or forms a direct contrast to what has been previously mentioned in the discourse (*contrastive focus*) (Chafe, 1976). Focus is prosodically encoded in many languages (Gussenhoven, 2007). In these languages, pitch and duration are identified as major prosodic cues for marking focus (Cooper et al., 1985; Heldner, 2003; Xu, 1999). How pitch and duration are used exactly is, however, language-specific. For example, many languages use both expanded pitch span and lengthened duration to distinguish narrow focus from non-focus and/or broad focus (Cooper et al. 1985 on English; Hanssen et al. 2018, Chen, under revision on Dutch; Xu 1999 on Standard Mandarin; Jannedy 2007 on Vietnamese), whereas some languages only use duration in focus-marking (Wang et al. 2011 on Deang; Wu & Xu 2010 on Cantonese).

It has been shown that cross-linguistic differences in prosodic systems shape the developmental trajectory of prosodic focus-marking in monolingual children (Arnhold, 2016; Chen, 2011; Grigos & Patel, 2010; Grünloh, Lieven, & Tomasello, 2015; Romøren, 2016; Yang, 2017; Yang & Chen, 2014). For example, Romøren (2016) found that Swedish-speaking monolingual children mastered the adult-like use of the phonological cue to focus earlier in Swedish (prominence H - a floating high tone that can be added to the lexical pitch pattern of the focal word) than in Dutch (i.e., accentuation). She suggested that the earlier mastery in Swedish-speaking children might be attributed to a more reliable mapping between prominence H and focus in Swedish (relative to the mapping between accentuation and focus in Dutch) (Romøren, 2016, p. 182). Using the same data elicitation method, Yang and Chen (2017) and Yang (2017) investigated the developmental trajectory of prosodic focus-marking in Mandarin-speaking and Seoul Korean-speaking monolingual children. They found that at the age of seven or eight years, the Seoul Korean-speaking children could vary duration but not pitch to distinguish narrow focus from post-focus, while Mandarin-speaking children could vary both duration and pitch to distinguish narrow focus from post- and pre-focus. Following Chen (2017), Yang (2017) suggested that the differences in the relative importance of phonological and phonetic cues to focus in Korean and Standard Mandarin. Specifically, in Korean, prosodic phrasing is considered the primary strategy and phonetic realisation of prosodic boundaries is the secondary strategy for encoding focus, while in Standard Mandarin, the phonetic manipulation of pitch and duration is the primary strategy for encoding focus.

4.1.2 Prosodic focus-marking in monolingual Mandarin-speaking adults and children

Research on prosodic focus-marking in adult native speakers of Mandarin has shown that focus is prosodically encoded via pitch and duration in Standard Mandarin (Chen, 2010; Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015; Xu, 1999; Yang, 2017; Yang & Chen, 2014). Regarding narrow focus, the pitch span and duration of its focal constituent are expanded and lengthened, relative to the non-focal

counterpart (Yang & Chen, 2014). Further, the pitch span of the post-focal constituent is compressed relative to its counterpart in broad focus condition in read speech (Xu, 1999). Regarding focus types that differed in the size of the focal constituent, both the pitch span and duration of the focal constituent are expanded and lengthened in narrow focus, relative to its counterpart in broad focus in read speech (Xu, 1999). However, in semi-spontaneous speech, only the duration of the focal constituent is lengthened in narrow focus, compared to its counterpart in broad focus (Yang, 2017; Yang & Chen, 2017). Regarding the focus types that differed in the contrastiveness of the focal constituent, both the pitch span and duration of the focal constituent are expanded and lengthened, in contrastive focus, relative to its counterpart in non-focus in read speech (Chen & Gussenhoven, 2008). However, in semi-spontaneous speech, neither pitch span nor duration is used to distinguish contrastive focus from narrow focus (Yang, 2017; Yang & Chen, 2017).

With respect to prosodic focus-marking in monolingual Mandarin-speaking children, Yang and Chen (2017) found that children use duration to distinguish narrow focus from non-focus (i.e., pre- and post-focus) in an adult-like way in all lexical tones at the age of four to five years, and use pitch-related cues for the same purpose only in some tones. At the age of seven to eight years, they can vary pitch span in an adult-like way to distinguish narrow focus from pre-focus in all lexical tones, and to distinguish narrow focus from post-focus in some tones (i.e., Tones 2 and 4). At the age of ten to eleven years, their use of pitch-related cues is not yet fully adult-like in certain tones. Furthermore, monolingual Mandarin-speaking children vary duration in an adult-like way in distinguishing narrow focus from broad focus at the age of four to five years. However, they also use pitch-related cues for this purpose, unlike monolingual Mandarin-speaking adults. At the age of seven to eight years, they no longer use pitch-related cues for this purpose, similar to monolingual Mandarin-speaking adults. Notably, at the age of ten to eleven years, they vary pitch span for distinguishing narrow focus from broad focus, different from monolingual Mandarin-speaking adults in semi-spontaneous speech, but similar to monolingual adult speakers in read speech (Xu, 1999). Finally, monolingual children do not vary

duration to distinguish contrastive focus and narrow focus, at the age of four to five years, unlike monolingual adults. At the age of seven to eight years, they vary neither pitch-related cues nor duration to distinguish narrow focus from contrastive focus, like monolingual Mandarin-speaking adults.

In contrast to children acquiring a West Germanic language, such as English (Wonnacott & Watson, 2008), German (Müller et al., 2006), and Dutch (Chen, 2009), Mandarin-speaking children acquire the use of duration earlier than pitch-related cues, Yang and Chen (2017) attributed the later acquisition of pitch-related cues in monolingual Mandarin-speaking children to the use of pitch for lexical purposes in Standard Mandarin. Specifically, they argued that it might be easier to master the use of duration relative to pitch as a prosodic cue for encoding focus, as there might be limited acoustic space available for focus-related manipulation in pitch in Standard Mandarin. Furthermore, Yang and Chen (2017) found that monolingual Mandarin-speaking children exhibit adult-like use of prosody in distinguishing focus types that differed in constituent size earlier than in distinguishing narrow focus and non-focus. This finding might be explained by the fact that only duration is used to distinguish focus types whereas both duration and pitch-related cues are used to a similar extent in distinguishing narrow focus from non-focus in Standard Mandarin (Yang, 2017).

4.1.3 The present study

In the present study, we examine prosodic focus-marking in Mandarin produced by Bai-Mandarin early bilingual children aged six to thirteen who have acquired Bai at home as their L1 and started their formal Standard Mandarin education at the age of six or seven. Further, they have been exposed to Standard Mandarin mainly via mass media and Dali Mandarin (a regional variety of Mandarin) informally in their daily life before the age of six. Bai is a Sino-Tibetan tone language spoken by more than one million Bai people, mainly in Dali Bai Autonomous Prefecture, Yunnan Province, China. As shown in Chapter 2, only duration is used in Bai for encoding focus, while both pitch and duration are used for focus-marking purposes in

Standard Mandarin (Chen, 2010; Ouyang & Kaiser, 2015; Xu, 1999; Yang, 2017; Yang & Chen, 2017). Specifically, only the duration of the focal constituent is lengthened in comparison to its counterpart in the post-focus for encoding narrow focus in Bai, while pitch span and duration of the focal constituent is expanded and lengthened in comparison to its non-focal counterpart in Standard Mandarin. Further, the duration of the focal constituent is lengthened in narrow focus relative to its counterpart in broad focus in Standard Mandarin (Yang & Chen, 2017). However, neither duration nor pitch-related cues are used for distinguishing focus types, i.e., narrow focus, broad focus and contrastive focus in Bai.

We aim to address the question as to what is the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children's Mandarin like. In particular, we investigate the use of prosody to encode narrow focus and focus types that differed in the size and contrastiveness of the focal constituent in different tones in Mandarin produced by Bai-Mandarin early bilingual children. Regarding the developmental route, considering that duration as a focus-marking cue might be easier to master relative to pitch due to the limited acoustic space available for focus-related manipulation in pitch for learners of tone languages, as suggested in Yang and Chen (Yang & Chen, 2017), and that Bai-Mandarin early bilingual children's L1 and L2 are both tone languages, we hypothesize that Bai-Mandarin early bilingual children will learn the use of duration earlier than the use of pitch-related cues for focus-marking purposes (Hypothesis a). Our prediction is then Bai-Mandarin early bilingual children will lengthen the duration of the focal constituent to distinguish narrow focus from pre- and post-focus in all lexical tones at an early stage of their Mandarin acquisition (i.e., six- to seven-year olds and/or nine- to ten-year olds), and use pitch-related cues for the same purpose only in some tones at an late stage of their Mandarin acquisition (i.e., nine- to ten-year-olds and/or twelve- to thirteen-year-olds). Further, considering that the earlier acquisition of using prosody in distinguishing focus types than in distinguishing narrow focus and non-focus is found in monolingual Mandarin-speaking children (Yang, 2017; Yang & Chen, 2017), and that early bilingual children's L2 acquisition can resemble monolingual

children's L1 acquisition in some linguistic domains (Dulay & Burt, 1973, 1974; Jia, 2003; Krashen, 1982; Paradis, 2005, 2007), we hypothesize that Bai-Mandarin early bilingual children will learn the use of prosody in distinguishing focus types earlier than in distinguishing narrow focus and non-focus (Hypothesis b). Our prediction is then that Bai-Mandarin early bilingual children will master the use of duration to distinguish narrow focus from broad focus in all lexical tones earlier than the use of duration and pitch-related cues for distinguishing narrow focus from non-focus in all lexical tones.

4.2 Method

4.2.1 Picture-matching game

A picture-matching game used in Yang and Chen (Yang & Chen, 2017, 2014) and Yang (2017) was adopted to elicit SVO sentences in Mandarin from Bai-Mandarin early bilingual children in a semi-spontanous setting. In this picture-matching game, three sets of pictures were used.

The experimenter and the participant each held a set of pictures ordered in a specific sequence; the third set of pictures was scattered around on a table. In the experimenter's pictures (the first set), there was always some information missing, e.g., a subject, an action (verb), an object, or all three. The participant's pictures (the second set) all contained a complete event. The participant's task was to help the experimenter with sorting out pictures from the first set (experimenter's pictures) and the third set (scattered around on the table) in matched pairs (see Figure 1).

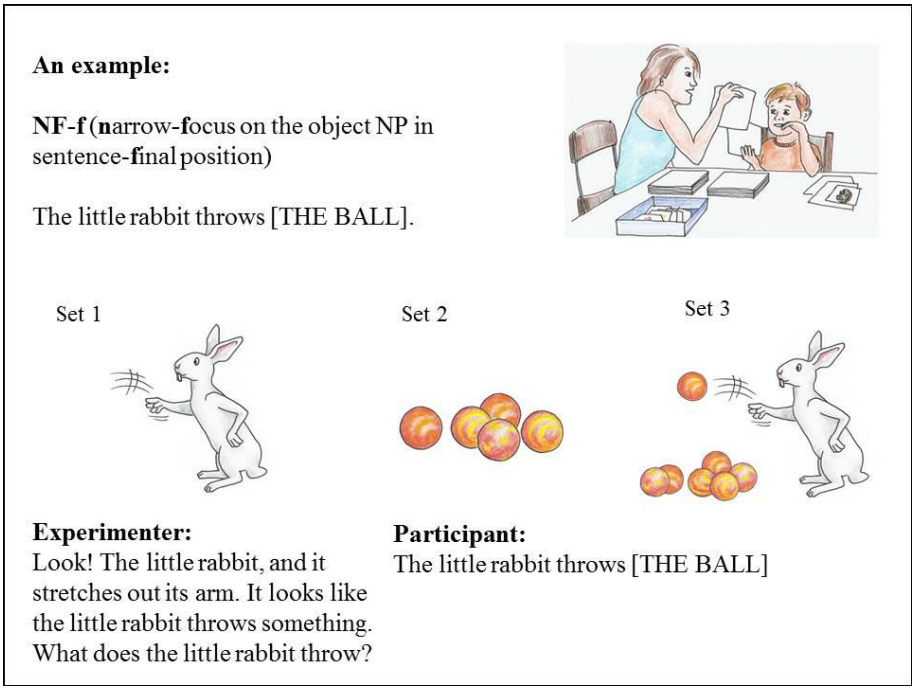

Figure 1. An example of a trial eliciting the sentence "The little rabbit throws [THE BALL]".

An example of a trial eliciting a target sentence with **n**arrow-focus on the object NP in sentence-**f**inal position (the NF-f condition): First, the experimenter took a picture (e.g., a rabbit with a waving arm) from her own set (set 1), drew the participant's attention to the picture and established what the picture was by saying, e.g., "Look! The rabbit, it waves its arm. It looks like the rabbit throws something." This was done to make sure that the entity in the picture was referentially given to the participant before the utterance of the question. Second, the experimenter asked a question about the picture (e.g., "What does the rabbit throw?"). Third, the participant took a picture from his or her set (set 2) and inspected it. The experimenter then repeated the question, followed by an answer from the participant (e.g., "The rabbit throws [THE BALL]."). Lastly, the experimenter found the picture containing the missing information in the third set (set 3) and paired it up with her

own picture. As to rules of the game, participants were explicitly asked to respond in full sentences and not to show their pictures to the experimenter. To ensure consistency in the participants' word choice, the picture-matching game was preceded by a picture-naming task, which was designed to familiarize the participants with the target words and the entities in the pictures used in the game.

4.2.2 Experimental materials

SVO sentences in five focus conditions were elicited via the picture-matching game: **n**arrow-**f**ocus on the subject NP in sentence-**i**nitial position (NF-i), **n**arrow-**f**ocus on the verb in sentence-**m**edial position (NF-m), **n**arrow-**f**ocus on the object NP in sentence-**f**inal position (NF-f), **b**road **f**ocus (BF) and **c**ontrastive **f**ocus on the verb in sentence-**m**edial position (CF-m). The focus condition was set up by a WH-question or a statement from the experimenter, as illustrated in examples (1) to (5).

(1) **NF-i**

Experimenter: 看！球。球飞在空中。看起来有小动物扔球。谁扔球？

Look! The ball. The ball is in the air. It looks like someone throws the ball. Who throws the ball?

Participant: [小熊]扔球。

[THE LITTLE BEAR] throws the ball.

(2) **NF-m**

Experimenter: 看！小狗，还有球。看起来小狗要弄球。小狗怎么弄球？

Look! The little dog, and the ball. It looks like the little dog does something to the ball. What does the little dog do to the ball?

Participant: 小狗[扔]球。

The little dog [THROWS] the ball.

(3) **NF-f**

Experimenter: 看！小兔。小兔的胳膊挥出去了。看起来小兔在扔东西。小兔扔什么？

Look! The little rabbit, and it stretches out its arm. It looks like the little rabbit throws something. What does the little rabbit throw?

Participant: 小兔扔[球]。

The little rabbit throws [THE BALL].

(4) **BF**

Experimenter: 看！阿姨什么都看不清。你的图片上讲了什么？

Look! My picture is very blurry. I cannot see anything clearly. What happens in your picture?

Participant: [小熊扔球]。

[THE LITTLE BEAR THROWS THE BALL].

(5) **CF-m**

Experimenter: 看！小猫，还有球。看起来小猫要弄球。我猜，小猫剪球。

Look! The little cat, and the ball. It looks like the little cat will do something to the ball. I will make a guess: The little cat [CUTS] the ball.

Participant: 小猫[扔]球。

The cat [THROWS] the ball.

The SVO sentences were constructed in such a way that each was a unique combination of a subject-noun and a VP (verb + object-noun). Each of the four lexical tones (i.e., high-level tone/Tone 1, rising tone/Tone 2, dipping tone/Tone 3, and falling tone/Tone 4) occurred in subject noun phrases, verbs, and object nouns. Four monosyllabic verbs and four monosyllabic object nouns formed 16 VPs, each

of which appeared in each focus condition (n = 5). This resulted in 80 VPs, and then four subject-noun phrases, which started with the word "xiǎo" (little)" and ended with four nouns carrying four lexical tones were evenly distributed over these 80 VPs to form 80 SVO target sentences. An overview of the words in SVO sentences is provided in Table 1.

Table 1. Overview of the words in SVO sentences, each word is presented in Chinese, Pinyin and English translation

	Subject	Verb	Object
Tone 1	小猫 xiǎo māo cat (little cat)	扔 rēng throw	书 shū book
Tone 2	小熊 xiǎo xióng bear (little bear)	埋 mái bury	球 qiú ball
Tone 3	小狗 xiǎo gǒu dog (little dog)	剪 jiǎn cut	笔 bǐ pen
Tone 4	小兔 xiǎo tù rabbit (little rabbit)	运 yùn transport	菜 cài vegetable

4.2.3 Participant and procedure

Twenty-five Bai-Mandarin bilingual children from three age groups took part in this study: eight six- to seven-year-olds (four girls and four boys, average age = 6.88, SD = 0.35), eight nine- to ten-year-olds (four girls and four boys, average age = 9.13, SD = 0.35) and nine twelve- to thirteen-year-olds (four girls and five boys, average age = 12.44, SD = 0.53). At the time of testing, the six- to seven-year-olds enrolled in a local primary school; the nine- to ten-year-olds started their third year of formal Standard Mandarin education; and the twelve- to thirteen-year-olds just started their sixth year of formal Standard Mandarin education (see Appendix D).

All the bilingual participants grew up in villages of Xizhou County, Dali Bai Autonomous Prefecture, Yunnan Province, China. They acquired Bai at home as their L1 and started their formal Standard Mandarin education at the age of six or seven. They were exposed to Mandarin mainly via mass media before the age of six. As they grew up in the same Bai-speaking community, came from families with similar socioeconomic status, attended the same local primary school, they formed a fairly homogeneous group. An overview of the participants is presented in Table 2.

Table 2. Overview of the Bai-Mandarin early bilingual child participants

Age group	AoA mean	AoA range	Length of formal Standard Mandarin education
6- to 7-year-olds	6;11	6;8-7;4	Just enrolled
9- to 10-year-olds	7;2	6;11-8;4	2 years
12- to 13-year-olds	7;5	6;11 -8;1	5 years

The participants were tested individually by a female experimenter, who was a native speaker (age = 27) of Standard Mandarin and only spoke Standard Mandarin with the participants. Each experiment was split into two sessions, each contained 40 trials. A break was introduced after one session to avoid loss of attention or fatigue in the participants. Each session took about 25-30 minutes for a participant to complete. All the test sessions were conducted in a quiet room in Jinhe Primary School, Xizhou County. The sessions were recorded using a portable ZOOM H1 digital recorder at a 44.1 kHz sampling rate and 16-bit accuracy. Each session was also video-recorded for future training purpose. After the experiment proper, a language background and home-usage questionnaire was sent to the children's caregivers.

4.2.4 Acoustic annotation

The auditory recordings from the participants were first orthographically annotated. An answer sentence was discarded for analysis if had one of the following features:

(1) it was not a response to the target question; (2) it contained self-corrections; (3) it contained hesitations (defined as a long "em" sound produced by the participants before answering questions); (4) it deviated from the target sentence in word choice or word order. In total, 72% of the obtained responses (n = 1433) were selected for the analysis. Specifically, 54% of the obtained responses from the six- to seven-year olds (n = 343), 80% of the obtained responses from the nine- to ten-year olds (n = 514), and 80% of the obtained responses from the twelve- to thirteen-year olds (n = 576) were selected for the analysis.

The **verbs** were the targets of prosodic analysis. They were focal constituents in the BF, NF-m, and CF-m conditions, pre-focal constituents in the NF-f condition, and post-focal constituents in the NF-i condition. This setup made it possible to investigate the effect of both focus and focus types on pitch and duration in the verbs. The verbs were acoustically annotated by examining the waveform, wide-band spectrum, and pitch track in Praat (Boersma & Weenink, 2006) in combination with auditory impressions (Turk et al., 2006). Two pitch-related landmarks and two segmental landmarks were labelled in each verb: pitch maximum, pitch minimum, word onset, and word offset.

The tonal targets of the lexical tones in Standard Mandarin were taken into consideration when the landmarks for pitch maximum and pitch minimum were inserted (Xu, 1997; Xu & Wang, 2001; Yang & Chen, 2014). Specifically, Tone 1 is a high-level tone, but we observed that it appeared as a slightly rising pitch contour and occasionally as a slightly falling pitch contour sentence-medially. The pitch-maximum was labelled after the pitch-minimum in the case of a rising Tone 1 syllable but before the pitch-minimum in the case of a falling Tone 1 syllable. Tone 2 is a rising tone, but we observed that it appeared as a fall-rise contour sentence-medially, as reported in Xu (1997) and Yang and Chen (2014). As the falling part is largely influenced by the preceding tone, the rising part is taken to contain the tonal targets of Tone 2 (Xu, 1997). Thus, the pitch-maximum was labelled after the pitch-minimum in the rising part of Tone 2. Similarly, Tone 4 is a falling tone, but we

observed that it appeared as a rise-fall contour sentence-medially, again, as reported in Xu (1997) and Yang and Chen (2014). As the rising part is subjected to the influence of the preceding tone, the falling part contains the tonal targets of Tone 4. Thus, the pitch-maximum was labelled before the pitch-minimum in the falling part of Tone 4. Tone 3 is a fall-rise tone when produced in isolation or sentence-finally, but it was realized as a falling pitch contour sentence-medially, and as a fall-rise contour when followed by another Tone 3, as reported in Xu (1997) and Yang and Chen (2014). We labelled the pitch-maximum before the pitch-minimum in the case of falling Tone 3 but after the pitch-minimum in the rising part of a falling-rising Tone 3.

The pitch values (in Hz) at the pitch-related landmarks, and time values (in seconds) at the segmental landmarks were automatically extracted by means of Praat scripts. Four measures were obtained from each verb: pitch maximum, pitch minimum, pitch span (i.e., the difference between the maximum pitch and the minimum pitch), and word duration (i.e., offset time minus onset time). In 112 of the selected responses (7.8%), an accurate measurement of pitch values was not possible. These responses were thus excluded from the analysis of pitch-related measurements.

4.3 Statistical analyses and results

4.3.1 Statistical analyses

In order to assess the effect of focus, we compared the measurements of the verbs between the unfocused conditions and the focused condition; that is, NF-m (verb is the focal constituent) vs. NF-i (verb is the post-focal constituent), and NF-m (verb is the focal constituent) vs. NF-f (verb is the pre-focal constituent). To investigate the effect of focus type that differed in the size of the focal constituent, we compared narrow focus on the verb (NF-m) with broad focus (BF). To investigate the effect of focus type that differed in the contrastiveness of the focal constituent, we compared contrastive focus on the verb (CF-m) with non-contrastive focus on the verb (NF-m).

Statistical analyses were conducted using mixed-effects modelling in R (R Core Team, 2014) with package "lme4" (Bates, Mächler, et al., 2015) and "lmerTest" (Kuznetsova et al. 2013). In all models, tone, focus conditions, and age were included as fixed factors, while speaker (i.e., the participants) and sentence (i.e., the answer sentences) were included as random factors. In each of the four pair-wise comparisons listed above, focus condition had two levels (the two focus conditions of interest), tone referred to the lexical tones of the verbs, which had four levels (Tones 1, 2, 3, and 4); and age had three levels (i.e., six- to seven-year-olds, nine- to ten-year-olds and twelve- to thirteen-year-olds). Outcome variables were the duration, pitch span, pitch maximum, and pitch minimum of the verbs.

Following Magezi (2015) and Field et al. (2012), our models were constructed and evaluated in a stepwise fashion. Specifically, we started from an intercept-only model, built the subsequent models by adding one new term/factor at a time, and then systematically compared the models which differed only by the presence or absence of one term. The comparison was done using a likelihood-ratio test, and the test statistic χ^2, degrees of freedom and *p*-value were reported for the added term. A *p*-value of less than 0.05 was considered to indicate that the added term contributed significantly to the model fit (Magezi, 2015). A detailed description of the model building is presented in Table 3.

Table 3. Model build-up procedure

Model	Added term
Model 0	only contains speaker, sentence as random intercepts
Model 1	+ tone
Model 2	+ focus condition
Model 3	+ tone : focus condition
Model 4	+ age
Model 5	+ age : focus condition
Model 6	+ age : tone
Model 7	+ focus condition: tone: age

When building the models, only the factors and interactions that significantly improved the fit of the model were retained until the best-fit model was determined. After the best-fit model was established, we only summarized and interpreted the model with the best fit. For each analysis, we reported the results of the model comparison first, and then the parameter estimates of the best-fit model. As we are primarily interested in the effect of focus on the outcome variables, we will concentrate on the main effect of focus or interactions involving the factor of focus. If the best-fit model contained the three-way interaction of focus, age and tone, we discuss how the speakers in each age group distinguished the two focus conditions in each tonal category by examining the interaction of focus and tone in each age group. If the best-fit model contained the two-way interaction of focus and age, we discuss how the speakers in each age group distinguished the two focus conditions by examining the main effect of focus in each age group. However, we do not discuss in detail the main effects of the factors when the interactions involving these factors are significant, two-way interactions when three-way interactions involving the same factors are significant, and interactions that do not involve the factors focus and age.

4.3.2 Duration

4.3.2.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The effect of focus (focus vs. post-focus) on duration was tested with the models specified in Table 3. The results of the fit and model comparisons are shown in Table 4. As can be seen, the best-fit model was the model containing the main effect of tone, χ^2 (3) = 12.167, $p < .01$, focus condition, χ^2 (1) = 9.469, $p < .01$, and a two-way interaction between age and tone, χ^2 (6) = 22.259, $p < .01$. Parameter estimates of the best-fit model are presented in Table 5. The main effect of focus condition was such that the duration of the verbs was significantly longer in narrow focus (224.2 ms, SD = 70.4) than in post-focus (205.7 ms, SD = 57), regardless of tone and age (b = 16.235, df = 32.2, t = 3.319, $p <$ .01), as shown in Figure 2.

Table 4. Duration, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−3242.7				
1 + tone	7	−3236.6	0 vs 1	12.167	3	.007**
2 + focus condition	8	−3231.9	1 vs 2	9.469	1	.002**
3 + tone : focus condition	11	−3230.4	2 vs 3	2.818	3	.421
4 + age	13	−3227.9	3 vs 4	5.068	2	.079.
5 + age : focus condition	15	−3226.1	4 vs 5	3.582	2	.167
6 + age : tone	21	−3215.0	5 vs 6	22.259	6	.001**
7 + focus condition : tone : age	27	−3210.8	6 vs 7	8.420	6	.209

Table 5. Duration, narrow focus (NF-m) vs. post-focus (NF-i), parameter estimates of the best-fit model

	Estimate	SE	*df*	*t* value	Pr (>|t|)
Fixed part					
Intercept	255.277	19.114	35.1	13.356	.000***
Tone 2	−4.626	10.464	150.0	−0.442	.659
Tone 3	−48.482	10.247	140.5	−4.731	.000***
Tone 4	3.085	10.424	147.7	0.296	.768
Narrow focus	16.235	4.892	32.2	3.319	.002**
9- to 10-year-olds	−42.358	25.829	29.5	−1.640	.112
12- to 13-year-olds	−63.048	25.156	29.8	−2.506	.018*
Tone 2: 9- to 10-year-olds	10.405	11.307	580.1	0.920	.358
Tone 3: 9- to 10-year-olds	30.203	11.116	579.4	2.717	.007**
Tone 4: 9- to 10-year-olds	−16.549	11.338	584.0	−1.460	.145
Tone 2: 12- to 13-year-olds	10.199	11.246	579.8	0.907	.365
Tone 3: 12- to 13-year-olds	30.662	10.962	580.0	2.797	.005**
Tone 4: 12- to 13-year-olds	−13.780	11.208	583.4	−1.229	.219
Random part	*Name*	S^2	*SE*		
Sentence	Intercept	119.2	10.92		
Speaker	Intercept	2382.6	48.81		
Residual		1376.7	37.10		

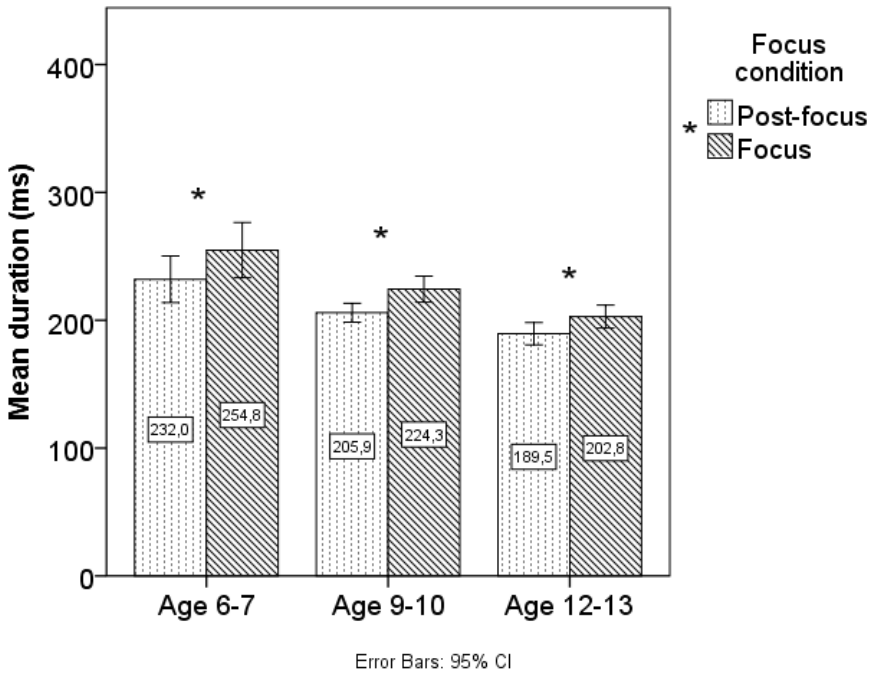

Figure 2. Mean duration (in ms) of post-focal constituent vs. focal constituent in three age groups. Significant differences are marked with an asterisk.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 6. As can be seen, the best-fit model was the model containing the main effect of tone, $\chi^2(3) = 20.381$, $p < .001$, focus condition, $\chi^2(1) = 5.5848$, $p < .05$, and a two-way interaction between age and tone, $\chi^2(6) = 35.339$, $p < .001$. Parameter estimates of the best-fit model are presented in Table 7. The main effect of focus condition was such that the duration of the verbs was significantly longer in narrow focus (224.2 ms, SD = 70.4) than in pre-focus (209.6 ms, SD = 65.1), regardless of tone and age ($b = 12.037$, $df = 31.4$, $t = 2.588$, $p < .05$), as shown in Figure 3.

Table 6. Duration, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−3179.1				
1 + tone	7	−3168.9	0 vs 1	20.381	3	.000***
2 + focus condition	8	−3166.1	1 vs 2	5.585	1	.018*
3 + tone : focus condition	11	−3164.8	2 vs 3	2.756	3	.431
4 + age	13	−3161.8	3 vs 4	5.897	2	.052.
5 + age : focus condition	15	−3160.9	4 vs 5	1.773	2	.412
6 + age : tone	21	−3143.2	5 vs 6	35.339	6	.000***
7 + focus condition : tone : age	27	−3140.5	6 vs 7	5.488	6	.483

Table 7. Duration, narrow focus (NF-m) vs. pre-focus (NF-f), parameter estimates of the best-fit model

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	268.760	19.633	35.3	13.689	.000***
Tone 2	3.006	10.651	177.8	0.282	.778
Tone 3	−66.216	10.532	174.5	−6.287	.000***
Tone 4	−6.360	10.395	165.4	−0.612	.541
Narrow focus	12.037	4.651	31.4	2.588	.015*
9- to 10-year-olds	−57.939	26.635	30.1	−2.175	.038*
12- to 13-year-olds	−77.855	25.906	30.2	−3.005	.005**
Tone 2: 9- to 10-year-olds	14.498	12.039	562.5	1.204	.229
Tone 3: 9- to 10-year-olds	48.927	12.073	560.9	4.053	.000**
Tone 4: 9- to 10- year-olds	−1.160	11.910	566.8	−0.097	.922
Tone 2: 12- to 13-year-olds	12.054	11.863	564.5	1.016	.310
Tone 3: 12- to 13-year-olds	56.484	11.672	561.3	4.839	.000***
Tone 4: 12- to 13- year-olds	−0.999	11.645	564.1	−0.086	.932
Random part	*Name*	S^2	*SE*		
Sentence	Intercept	90.82	9.53		
Speaker	Intercept	2507.81	50.08		
Residual		1510.00	38.86		

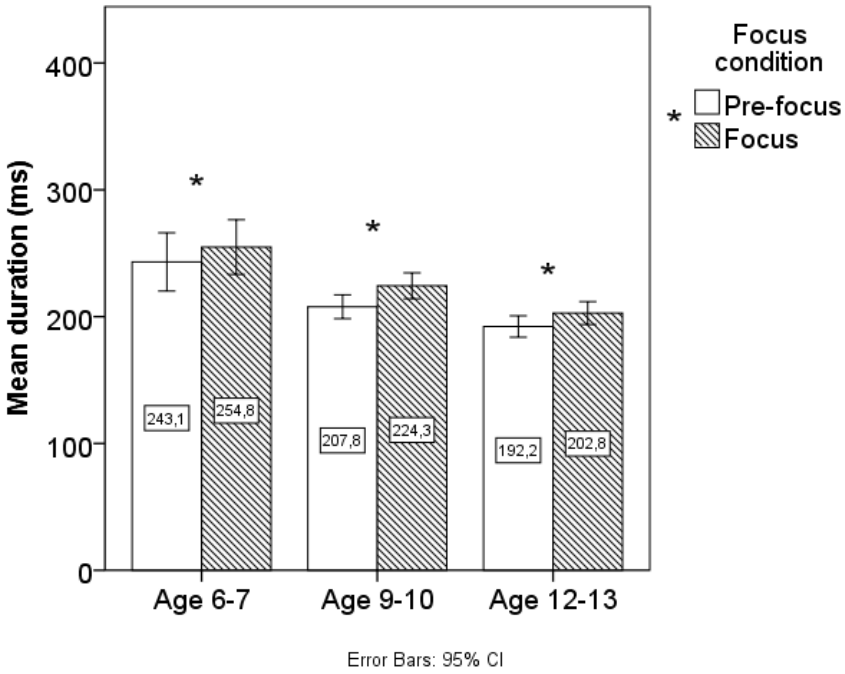

Figure 3. Mean duration (in ms) of pre-focal constituent vs. focal constituent in three age groups. Significant differences are marked with an asterisk.

4.3.2.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 8. As can be seen, the best-fit model was the model (model 7) containing a three-way interaction between tone, focus condition and age, χ^2 (6) = 13.124, $p < .05$. Parameter estimates of the best-fit model are presented in Table 9.

We explored the details of the three-way interaction between tone, focus condition and age by examining the interaction between focus condition and tone in each age group. In the six- to seven-year-olds, subsequent analysis revealed no significant main effect of focus condition (p = .396) or significant interaction between focus condition and tone (p = .133), indicating that they did not use duration to distinguish narrow focus from broad focus, regardless of tone. In the nine- to ten-year-olds,

there was no significant main effect of focus condition (p = .068) or significant interaction between focus condition and tone (p = .595), indicating that they did not use duration to distinguish narrow focus from broad focus, regardless of tone. In the twelve- to thirteen-year-olds, there was no significant main effect of focus condition (p = .658) or significant interaction between focus condition and tone (p = .722), indicating that they did not use duration to distinguish narrow focus from broad focus, regardless of tone. We then explored the details of the three-way interaction between tone, focus condition and age by examining the interaction between age and tone in each focus condition. Subsequent analysis revealed a significant interaction between age and tone in narrow focus (p < .001), but not in broad focus (p = .803).

There was thus no evidence that the bilingual children made use of duration to distinguish narrow focus from broad focus, regardless of tone and age. The three-way interaction between tone, focus condition and age is visualized in Figure 4.

Table 8. Duration, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−2867.7				
1 + tone	7	−2858.6	0 vs 1	18.078	3	.000***
2 + focus condition	8	−2858.2	1 vs 2	0.889	1	.346
3 + tone : focus condition	11	−2857.9	2 vs 3	0.499	3	.919
4 + age	13	−2854.6	3 vs 4	6.575	2	.037*
5 + age : focus condition	15	−2853.3	4 vs 5	2.745	2	.254
6 + age : tone	21	−2846.6	5 vs 6	13.267	6	.039*
7 + focus condition : tone : age	27	−2840.1	6 vs 7	13.124	6	.041*

Table 9. Duration, narrow focus (NF-m) vs. broad focus (BF), parameter estimates of the best-fit model (Model 7)

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	269.838	22.118	47.8	12.2	.000***
Tone 2	17.387	16.851	212.2	1.032	.303
Tone 3	−11.252	17.544	237.5	−0.641	.522
Tone 4	3.852	19.8	311.4	0.195	.846
Narrow focus	15.738	16.928	215.5	0.930	.354
9- to 10-year-olds	−34.54	29.698	39.4	−1.163	.252
12- to 13-year-olds	−63.92	28.806	39.1	−2.219	.032*
Tone 2: narrow focus	−27.385	22.744	186.5	−1.204	.230
Tone 3: narrow focus	−63.502	23.279	199.7	−2.728	.007**
Tone 4: narrow focus	−3.234	24.942	241.2	−0.130	.897
Narrow focus:9- to 10-year-olds	−23.612	19.273	504.6	−1.225	.221
Narrow focus:12- to 13-year-olds	−17.623	18.751	501.4	−0.940	.348
Tone 2: 9- to 10-year-olds	−6.969	19.503	502.5	−0.357	.721
Tone 3: 9- to 10-year-olds	−6.892	20.157	504.0	−0.342	.733
Tone 4: 9- to 10-year-olds	8.394	23.185	506.7	0.362	.717
Tone 2: 12- to 13-year-olds	−11.732	18.9	499.5	−0.621	.535
Tone 3: 12- to 13-year-olds	−16.705	19.644	499.0	−0.850	.396
Tone 4: 12- to 13-year-olds	−11.746	21.981	503.9	−0.534	.593
Tone 2:narrow focus:9- to 10-year-olds	28.438	25.987	501.3	1.094	.274
Tone 3:narrow focus:9- to 10-year-olds	63.904	26.582	501.5	2.404	.017*
Tone 4:narrow focus:9- to 10-year-olds	−16.172	28.898	504.7	−0.560	.576
Tone 2:narrow focus:12- to 13-year-olds	32.129	25.472	498.7	1.261	.208
Tone 3:narrow focus:12- to 13-year-olds	75.556	25.950	499.4	2.912	.004**
Tone 4:narrow focus:12- to 13-year-olds	1.442	27.7	501.7	0.052	.959
Random part	*Name*	S^2	*SE*		
Sentence	Intercept	94.94	9.744		

Speaker	Intercept	2697.23	51.935
Residual		1574.52	39.680

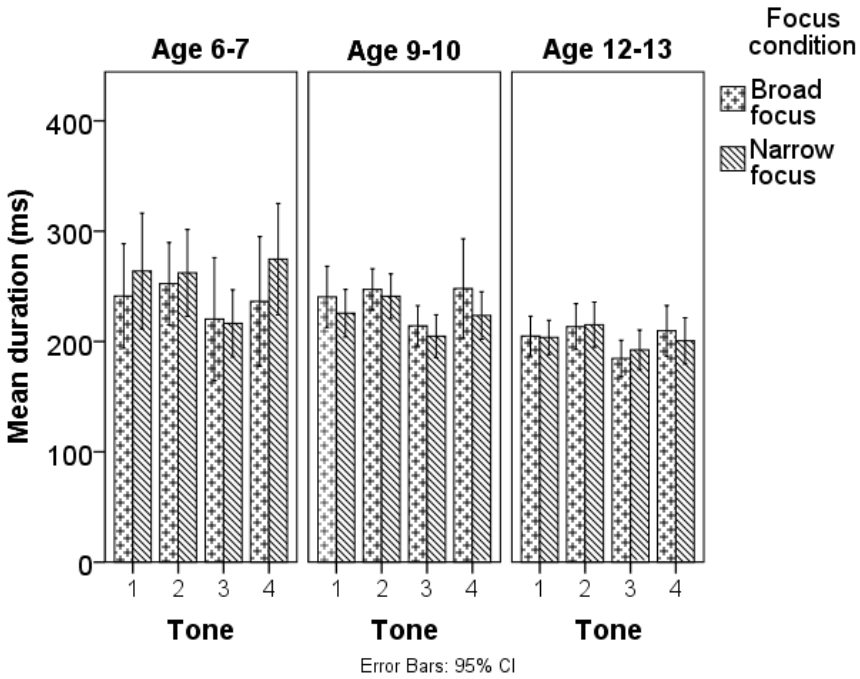

Figure 4. Mean duration (in ms) of broad focus vs. narrow focus in in four lexical tones and three age groups. Significant differences are marked with an asterisk.

4.3.2.3 Effect of contrastiveness: contrastive focus (CF-m) vs. non-contrastive focus (NF-m)

The results of the fit and model comparisons are shown in Table 10. As can be seen, the best-fit model was the model (model 7) containing a three-way interaction between tone, focus condition and age, χ^2 (6) = 21.14, p < .01. Parameter estimates of the best-fit model are presented in Table 11.

We explored the details of the three-way interaction between tone, focus condition and age by examining the interaction between focus condition and tone in each age

group. In the six- to seven-year-olds, subsequent analysis revealed no significant main effect of focus condition (p = .714), but a significant interaction between focus condition and tone (p < .05). We observed that the Tone 4 verbs in the narrow focus condition were averagely produced with much longer duration (36.1 ms) than their counterparts in the contrastive focus condition, which was different from Tone 1, 2 and 3 verbs. We then explored the details of the two-way interaction between focus condition and tone by examining the effect of focus condition in each tone in six- to seven-year-olds. Subsequent analysis revealed that no significant main effect of focus condition in Tone 1 (p = .922), Tone 2 (p = .228), or Tone 3 (p = .152), although the main effect of focus condition in Tone 4 was approaching significance (p = .099) in six- to seven-year-olds. These analyses indicated that the six- to seven-year-olds did not use duration to distinguish narrow focus from contrastive focus, regardless of tone. In the nine- to ten-year-olds, subsequent analysis revealed no significant main effect of focus condition (p = .263) or significant interaction of focus condition and tone (p = .777), indicating that they did not use duration to distinguish narrow focus from contrastive focus, regardless of tone. In the twelve- to thirteen-year-olds, subsequent analysis revealed no significant main effect of focus condition (p = .203) or significant interaction of focus condition and tone (p = .459), indicating that they did not use duration to distinguish narrow focus from contrastive focus, regardless of tone.

There was thus no evidence that the bilingual children made use of duration to distinguish narrow focus from contrastive focus, regardless of tone and age. The three-way interaction between tone, focus condition and age is illustrated in Figure 5.

Table 10. Duration, narrow focus (NF-m) vs. contrastive focus (CF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−3133.0				
1 + tone	7	−3124.6	0 vs 1	16.74	3	.000***
2 + focus condition	8	−3123.9	1 vs 2	1.504	1	.220
3 + tone : focus condition	11	−3123.3	2 vs 3	1.046	3	.79
4 + age	13	−3120.8	3 vs 4	5.161	2	.076.
5 + age : focus condition	15	−3120.7	4 vs 5	0.066	2	.967
6 + age : tone	21	−3114.9	5 vs 6	11.727	6	.068.
7 + focus condition : tone : age	27	−3104.3	6 vs 7	21.14	6	.002**

Table 11. Duration, narrow focus (NF-m) vs. contrastive focus (CF-m), parameter estimates of the best-fit model (Model 7)

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	280.254	21.531	46.9	13.016	.000***
Tone 2	12.601	16.318	202.6	0.772	.441
Tone 3	−36.576	16.155	189.1	−2.264	.025*
Tone 4	−34.716	15.168	159.8	−2.289	.023*
Narrow focus	−2.723	16.482	206.0	−0.165	.869
9- to 10-year-olds	−40.877	28.853	38.7	−1.417	.165
12- to 13-year-olds	−67.460	28.036	38.7	−2.406	.021*
Tone 2: narrow focus	−22.902	22.638	190.2	−1.012	.313
Tone 3: narrow focus	−35.999	22.567	183.8	−1.595	.112
Tone 4: narrow focus	39.473	21.469	160.5	1.839	.068.
Narrow focus:9- to 10-year-olds	−9.896	18.793	546.4	−0.527	.599
Narrow focus:12- to 13-year-olds	−6.121	18.356	542.3	−0.333	.739
Tone 2: 9- to 10-year-olds	−16.131	18.801	541.8	−0.858	.391
Tone 3: 9- to 10-year-olds	13.051	19.086	547.9	0.684	.494
Tone 4: 9- to 10-year-olds	23.191	17.815	543.4	1.302	.194
Tone 2: 12- to 13-year-olds	−12.901	18.601	542.1	−0.694	.488
Tone 3: 12- to 13-year-olds	12.421	18.367	543.5	0.676	.499
Tone 4: 12- to 13-year-olds	36.321	17.328	537.8	2.096	.037*
Tone 2:narrow focus:9- to 10-year-olds	38.240	25.873	544.1	1.478	.140
Tone 3:narrow focus:9- to 10-year-olds	41.606	26.242	547.7	1.585	.113
Tone 4:narrow focus:9- to 10-year-olds	−35.829	24.996	545.8	−1.433	.152
Tone 2:narrow focus:12- to 13-year-olds	34.034	25.658	543.4	1.326	.185
Tone 3:narrow focus:12- to 13-year-olds	44.068	25.457	546.3	1.731	.084.
Tone 4:narrow focus:12- to 13-year-olds	−51.241	24.394	542.4	−2.101	.036*
Random part	*Name*	S^2	*SE*		
Sentence	Intercept	92.14	9.599		

Speaker	Intercept	2569.88	50.694
Residual		1683.68	41.033

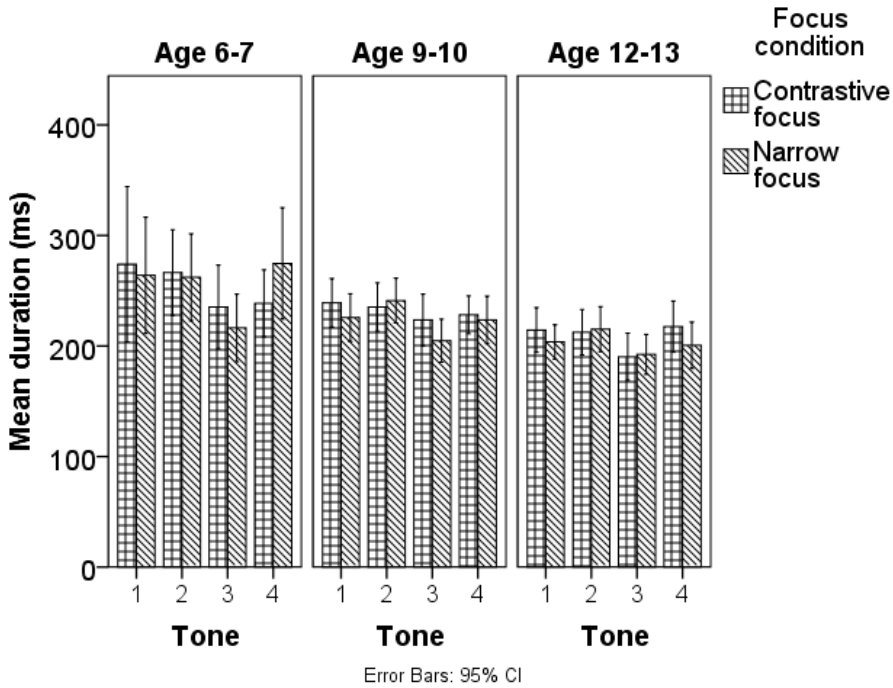

Figure 5. Mean duration (in ms) of contrastive focus vs. narrow focus (non-contrastive focus) in four lexical tones and three age groups. Significant differences are marked with an asterisk.

4.3.3 Pitch span

4.3.3.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 12. As can be seen, the best-fit model was the model (model 7) containing a three-way interaction between tone, focus condition

and age, χ^2 (6) = 13.009, $p < .05$. Parameter estimates of the best-fit model are presented in Table 13.

We explored the details of the three-way interaction between tone, focus condition and age by examining the interaction between focus condition and tone in each age group. In the six- to seven-year-olds, subsequent analysis revealed no significant main effect of focus condition ($p = .669$) or significant interaction between focus condition and tone ($p = .246$), indicating that they did not use pitch span to distinguish narrow focus from post-focus, regardless of tone. In the nine- to ten-year-olds, there was no significant main effect of focus ($p = .371$) or significant interaction of focus and tone ($p = .319$), indicating that they did not use pitch span to distinguish narrow focus from post-focus, regardless of tone. In the twelve- to thirteen-year-olds, there was no significant main effect of focus ($p = .710$) or significant interaction of focus and tone ($p = .784$), indicating that they did not use pitch span to distinguish narrow focus from post-focus, regardless of tone. We then explored the details of the three-way interaction between tone, focus condition and age by examining the interaction between age and tone in each focus condition. Subsequent analysis revealed a significant interaction of age and tone in post-focus ($p < .01$), but not in broad focus ($p = .175$).

There was thus no evidence that the bilingual children made use of pitch span to distinguish narrow focus from post-focus, regardless of tone and age. The three-way interaction between tone, focus condition and age is illustrated in Figure 6.

Table 12. Pitch span, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2595.4				
1 + tone	7	−2581.9	0 vs 1	27.08	3	.000***
2 + focus condition	8	−2581.7	1 vs 2	0.232	1	.630
3 + tone : focus condition	11	−2581.1	2 vs 3	1.232	3	.745
4 + age	13	−2574.1	3 vs 4	14.126	2	.000***
5 + age : focus condition	15	−2573.4	4 vs 5	1.409	2	.494
6 + age : tone	21	−2564.7	5 vs 6	17.397	6	.008**
7 + focus condition : tone : age	27	−2558.2	6 vs 7	13.009	6	.043*

Table 13. Pitch span, narrow focus (NF-m) vs. post-focus (NF-i), parameter estimates of the best-fit model (Model 7)

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	37.088	6.787	108.7	5.464	.000***
Tone 2	3.139	9.084	129.5	0.346	.730
Tone 3	41.318	8.402	98.3	4.918	.000***
Tone 4	6.339	8.849	115.6	0.716	.475
Narrow focus	−5.170	8.710	112.1	−0.656	.513
9- to 10-year-olds	−13.169	7.486	112.7	−1.759	.081.
12- to 13-year-olds	−16.146	7.449	122.0	−2.168	.032*
Tone 2: narrow focus	1.709	12.579	120.4	0.136	.892
Tone 3: narrow focus	−7.930	12.131	106.2	−0.654	.515
Tone 4: narrow focus	22.680	12.203	107.2	1.859	.066.
Narrow focus:9- to 10-year-olds	2.786	8.652	523.6	0.322	.748
Narrow focus:12- to 13-year-olds	9.322	8.610	523.5	1.083	.279
Tone 2: 9- to 10-year-olds	−1.184	9.264	525.7	−0.128	.898
Tone 3: 9- to 10-year-olds	−23.177	8.387	524.9	−2.763	.006**
Tone 4: 9- to 10-year-olds	4.325	8.836	537.1	0.489	.625
Tone 2: 12- to 13-year-olds	−10.982	9.096	525.8	−1.207	.228
Tone 3: 12- to 13-year-olds	−23.770	8.432	526.2	−2.819	.005**
Tone 4: 12- to 13-year-olds	11.067	8.928	536.5	1.240	.216
Tone 2:narrow focus:9- to 10-year-olds	−1.512	12.732	526.4	−0.119	.906
Tone 3:narrow focus:9- to 10-year-olds	19.138	12.151	523.6	1.575	.116
Tone 4:narrow focus:9- to 10-year-olds	−8.898	12.194	525.9	−0.730	.466
Tone 2:narrow focus:12- to 13-year-olds	−0.537	12.666	527.2	−0.042	.966
Tone 3:narrow focus:12- to 13-year-olds	2.151	12.091	524.7	0.178	.859
Tone 4:narrow focus:12- to 13-year-olds	−27.927	12.196	525.1	−2.290	.022*
Random part	*Name*	S^2	*SE*		
Sentence	Intercept	56.23	7.518		

Speaker	Intercept	80.38	8.966
Residual		388.56	19.712

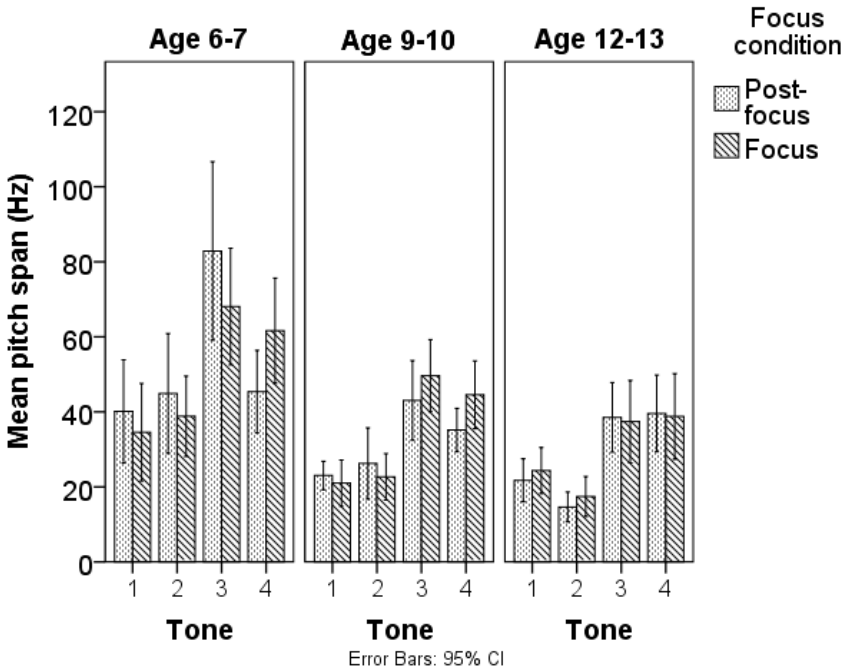

Figure 6. Mean pitch span (in Hz) of focus vs. post-focus in four lexical tones and three age groups. Significant differences are marked with an asterisk.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 14. As can be seen, the best-fit model was the model containing the main effects of tone, $\chi^2(3) = 28.047$, $p < .001$, and age, $\chi^2(2) = 11.73$, $p < .01$. There was thus no evidence that Bai-Mandarin early bilingual children varied pitch span in their Mandarin to distinguish narrow focus from pre-focus, regardless of tone and age.

Table 14. Pitch span, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2471.8				
1 + tone	7	−2457.8	0 vs 1	28.047	3	.000***
2 + focus condition	8	−2457.8	1 vs 2	0.005	1	.943
3 + tone : focus condition	11	−2457.7	2 vs 3	0.186	3	.980
4 + age	13	−2451.9	3 vs 4	11.73	2	.003**
5 + age : focus condition	15	−2451.2	4 vs 5	1.259	2	.533
6 + age : tone	21	−2445.8	5 vs 6	10.819	6	.094.
7 + focus condition : tone : age	27	−2443.2	6 vs 7	5.291	6	.507

4.3.3.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 15. As can be seen, the best-fit model was the model containing the main effect of tone, $\chi^2(3) = 31.633$, $p < .001$, and age, $\chi^2(2) = 11.708$, $p < .01$. There was thus no evidence that Bai-Mandarin early bilingual children varied pitch span in their Mandarin to distinguish narrow focus from broad focus, regardless of tone and age.

Table 15. Pitch span, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2291.1				
1 + tone	7	−2275.3	0 vs 1	31.633	3	.000***
2 + focus condition	8	−2275.3	1 vs 2	0.001	1	.973
3 + tone : focus condition	11	−2275.0	2 vs 3	0.525	3	.913
4 + age	13	−2269.2	3 vs 4	11.708	2	.003**
5 + age : focus condition	15	−2269.0	4 vs 5	0.235	2	.889
6 + age : tone	21	−2266.6	5 vs 6	4.816	6	.568
7 + focus condition : tone : age	27	−2262.2	6 vs 7	8.954	6	.176

4.3.3.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 16. As can be seen, the best-fit model obtained was the model containing a two-way interaction between tone and age, $\chi^2(6) = 15.072$, $p < .05$. There was thus no evidence that Bai-Mandarin bilingual children varied pitch span in their Mandarin to distinguish contrastive focus from narrow focus, regardless of tone and age.

Table 16. Pitch span, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2508.5				
1 + tone	7	−2494.0	0 vs 1	28.981	3	.000***
2 + focus condition	8	−2493.6	1 vs 2	0.773	1	.379
3 + tone : focus condition	11	−2493.2	2 vs 3	0.931	3	.818
4 + age	13	−2488.2	3 vs 4	9.920	2	.007**
5 + age : focus condition	15	−2487.4	4 vs 5	1.651	2	.438
6 + age : tone	21	−2479.8	5 vs 6	15.072	6	.020*
7 + focus condition : tone : age	27	−2474.1	6 vs 7	11.442	6	.076.

4.3.4 Pitch maximum

4.3.4.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 17. As can be seen, the best-fit model obtained was the model containing a two-way interaction between tone and age, χ^2 (6) = 15.108, $p < .05$. There was thus no evidence that Bai-Mandarin bilingual children varied pitch maximum in their Mandarin to distinguish narrow focus from post-focus, regardless of tone and age.

Table 17. Pitch maximum, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2719.9				
1 + tone	7	−2687.6	0 vs 1	64.679	3	.000***
2 + focus condition	8	−2687.0	1 vs 2	1.099	1	.294
3 + tone : focus condition	11	−2684.4	2 vs 3	5.102	3	.165
4 + age	13	−2679.6	3 vs 4	9.716	2	.008**
5 + age : focus condition	15	−2679.5	4 vs 5	0.163	2	.922
6 + age : tone	21	−2672.0	5 vs 6	15.108	6	.019*
7 + focus condition : tone : age	27	−2666.1	6 vs 7	11.636	6	.071.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 18. As can be seen, the best-fit model obtained was the model containing the interaction between tone and age, χ^2 (6) = 44.028, $p < .001$. There was thus no evidence that Bai-Mandarin bilingual children varied pitch maximum in their Mandarin to distinguish narrow focus from pre-focus, regardless of tone and age.

Table 18. Pitch maximum, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2577.2				
1 + tone	7	−2536.8	0 vs 1	80.836	3	.000***
2 + focus condition	8	−2536.8	1 vs 2	0.006	1	.941
3 + tone : focus condition	11	−2536.1	2 vs 3	1.341	3	.720
4 + age	13	−2531.0	3 vs 4	10.204	2	.006**
5 + age : focus condition	15	−2531	4 vs 5	0.003	2	.999
6 + age : tone	21	−2509	5 vs 6	44.028	6	.000 ***
7 + focus condition : tone : age	27	−2508.2	6 vs 7	1.643	6	.950

4.3.4.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 19. As can be seen, the best-fit model was the model containing a two-way interaction between tone and age, χ^2 (6) = 34.076, $p < .001$. Thus, there was thus no evidence that Bai-Mandarin bilingual children varied pitch maximum in their Mandarin to distinguish narrow focus from broad focus, regardless of tone and age.

Table 19. Pitch maximum, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2338.5				
1 + tone	7	−2300.8	0 vs 1	75.384	3	.000***
2 + focus condition	8	−2300.8	1 vs 2	0.010	1	.921
3 + tone : focus condition	11	−2300.6	2 vs 3	0.442	3	.931
4 + age	13	−2295.7	3 vs 4	9.786	2	.007**
5 + age : focus condition	15	−2294.9	4 vs 5	1.535	2	.464
6 + age : tone	21	−2277.9	5 vs 6	34.076	6	.000***
7 + focus condition : tone : age	27	−2276.6	6 vs 7	2.682	6	.848

4.3.4.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 20. As can be seen, the best-fit model was the model containing a two-way interaction between tone and age, χ^2 (6) = 35.757, p < .001. There was thus no evidence that Bai-Mandarin bilingual children varied pitch maximum in their Mandarin to distinguish contrastive focus from narrow focus (i.e., non-contrastive focus), regardless of tone and age.

Table 20. Pitch maximum, contrastive focus (NF-m) vs. narrow focus (CF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2588.3				
1 + tone	7	−2550.5	0 vs 1	75.682	3	.000***
2 + focus condition	8	−2550.3	1 vs 2	0.333	1	.564
3 + tone : focus condition	11	−2550.1	2 vs 3	0.345	3	.951
4 + age	13	−2545.1	3 vs 4	10.056	2	.007**
5 + age : focus condition	15	−2545.0	4 vs 5	0.159	2	.923
6 + age : tone	21	−2527.2	5 vs 6	35.757	6	.000***
7 + focus condition : tone : age	27	−2524.9	6 vs 7	4.503	6	.609

4.3.5 Pitch minimum

4.3.5.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 21. As can be seen, the best-fit model was the model only containing the main effect of tone, χ^2 (3) = 72.167, $p < .001$. There was thus no evidence that Bai-Mandarin bilingual children varied pitch minimum in their Mandarin to distinguish narrow focus from post-focus, regardless of tone and age.

Table 21. Pitch minimum, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2704.0				
1 + tone	7	−2667.9	0 vs 1	72.167	3	.000***
2 + focus condition	8	−2667.5	1 vs 2	0.771	1	.380
3 + tone : focus condition	11	−2664.9	2 vs 3	5.075	3	.166
4 + age	13	−2663.0	3 vs 4	3.985	2	.136
5 + age : focus condition	15	−2662.1	4 vs 5	1.740	2	.419
6 + age : tone	21	−2658.7	5 vs 6	6.772	6	.343
7 + focus condition : tone : age	27	−2652.9	6 vs 7	11.578	6	.072.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 22. As can be seen, the best-fit model was the model containing a two-way interaction between tone and age, χ^2 (6) = 35.248, $p < .001$. There was thus no evidence that Bai-Mandarin bilingual children varied pitch minimum in their Mandarin to distinguish narrow focus from pre-focus, regardless of tone and age.

Table 22. Pitch minimum, narrow focus (NF-m) vs. pre-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2523.2				
1 + tone	7	−2478.4	0 vs 1	89.609	3	.000***
2 + focus condition	8	−2478.2	1 vs 2	0.333	1	.564
3 + tone : focus condition	11	−2477.3	2 vs 3	1.745	3	.627
4 + age	13	−2475.5	3 vs 4	3.671	2	.160
5 + age : focus condition	15	−2475.4	4 vs 5	0.198	2	.906
6 + age : tone	21	−2457.8	5 vs 6	35.248	6	.000***
7 + focus condition : tone : age	27	−2456.4	6 vs 7	2.757	6	.839

4.3.5.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 23. As can be seen, the best-fit model was the model containing a two-way interaction between tone and age, χ^2 (6) = 29.281, p < .001. There was thus no evidence that Bai-Mandarin bilingual children varied pitch minimum in their Mandarin to distinguish narrow focus from broad focus, regardless of tone and age.

Table 23. Pitch minimum, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2295.7				
1 + tone	7	−2250.1	0 vs 1	91.134	3	.000***
2 + focus condition	8	−2250.1	1 vs 2	0.005	1	.945
3 + tone : focus condition	11	−2249.9	2 vs 3	0.504	3	.918
4 + age	13	−2247.7	3 vs 4	4.418	2	.110
5 + age : focus condition	15	−2247.4	4 vs 5	0.543	2	.762
6 + age : tone	21	−2232.8	5 vs 6	29.281	6	.000***
7 + focus condition : tone : age	27	−2232.1	6 vs 7	1.323	6	.970

4.3.5.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 24. As can be seen, the best-fit model was the model containing a two-way interaction between tone and age, χ^2 (6) = 47.419, p < .001. There was thus no evidence that Bai-Mandarin bilingual children varied pitch minimum in their Mandarin to distinguish contrastive focus from narrow focus (i.e., non-contrastive focus), regardless of tone and age.

Table 24. Pitch minimum, contrastive focus (NF-m) vs. narrow focus (CF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−2543.4				
1 + tone	7	−2498.2	0 vs 1	90.292	3	.000***
2 + focus condition	8	−2498.2	1 vs 2	0.074	1	.785
3 + tone : focus condition	11	−2498.1	2 vs 3	0.275	3	.965
4 + age	13	−2496.3	3 vs 4	3.569	2	.168
5 + age : focus condition	15	−2495.9	4 vs 5	0.852	2	.653
6 + age : tone	21	−2472.2	5 vs 6	47.419	6	.000***
7 + focus condition : tone : age	27	−2466.8	6 vs 7	10.825	6	.094.

4.3.6 Interim summary

First, we found that the bilingual children acquired the use of duration to encode narrow focus at the age of six to seven years. Specifically, all the age groups lengthened the duration of the focal constituent in comparison to the pre-focal and post-focal constituent for encoding narrow focus in their L2-Mandarin, like monolingual Mandarin-speaking adults. However, we found that none of the age groups used the pitch-related prosodic cues, including pitch span, pitch maximum, and pitch minimum to differentiating narrow focus from non-focus in any of the lexical tones. Furthermore, the bilingual children varied neither duration nor pitch-related prosodic cues to differentiate narrow focus from broad focus, regardless of tone and age. Finally, they made no use of duration or pitch-related prosodic cues for distinguishing contrastive focus from non-contrastive focus, regardless of tone and age.

4.4 Discussion and conclusion

To illustrate the similarities and differences in the developmental trajectory of prosodic focus-marking in Mandarin in Bai-Mandarin early bilingual children and monolingual Mandarin-speaking children, we present an overview of the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children (present study) and monolingual Mandarin-speaking children (Yang & Chen, 2017) in Table 25. These two studies employed the same data elicitation method and are thus ideal for comparison.

As shown in Table 25, there are both similarities and differences in the route and rate of acquisition of prosodic focus-marking between Bai-Mandarin early bilingual children's L2 and Mandarin-speaking monolingual children's L1. Regarding the developmental route, Bai-Mandarin bilingual children are similar to monolingual Mandarin-speaking children in their earlier mastery of duration for focus-marking purposes relative to pitch in their L2-Mandarin. Specifically, monolingual Mandarin-speaking children master the use of duration for encoding narrow focus in a Mandarin-speaking adult-like way in all the lexical tones at the age of four to five years, while the use of pitch-related cues for the same purpose is only present in some tones. Similarly, when Bai-Mandarin bilingual children can use duration for distinguishing narrow focus from non-focus (i.e., pre- and post-focus) in their Mandarin at the age of six to seven years, they cannot use pitch-related cues in a native-like manner. Thus, our results confirm Hypothesis a, stating that Bai-Mandarin bilingual children acquire the use of duration for focus-marking purposes earlier relative to the use of pitch.

Our finding of the earlier acquisition of duration as a prosodic cue for focus-marking purposes relative to pitch is compatible with the assumption that it is easier to master the use of duration relative to pitch as a prosodic cue for encoding focus due to the limited acoustic space available for focus-related manipulation in pitch for learners of tone languages at the early stage of language acquisition, as suggested in Yang and Chen (2017) and Yang (2017). However, since Bai-Mandarin early bilinguals'

L1, Bai, also exploits duration for encoding focus, it is also possible that the bilinguals benefit from positive L1 transfer in learning the use of duration in their L2-Mandarin, as suggested in the previous studies of Tsat-Mandarin (Wang et al., 2012), Italian-Spanish (Barnes & Michnowicz, 2015) and Quechua-Spanish (Muntendam & Torreira, 2016; Van Rijswijk & Muntendam, 2012) bilinguals. In addition, exposure to Dali Mandarin which only uses duration for prosodic focus-marking (Chapter 3) in the first six years of life may also contribute to bilinguals' ealy use of duration in focus-marking in Mandarin. Together, L1 transfer and possible influence of Dali Mandarin may explain why Bai-Mandarin bilinguals can master the use of duration for encoding focus at the age of six to seven years when they just started their formal Mandarin education (more on this in Chapter 6).

However, our results provide no evidence for Hypothesis b, stating that Bai-Mandarin early bilingual children will learn the use of prosody in distinguishing focus types earlier than in distinguishing narrow focus and non-focus, similar to monolingual Mandarin-speaking children. We have found that Bai-Mandarin bilingual children vary neither duration nor pitch in distinguishing focus types at the age of six to seven years. However, at the same age, they use duration for distinguishing narrow focus from non-focus (i.e., pre- and post-focus) in a native-like way in their Mandarin. This is rather unexpected, especially given that duration is used to distinguish focus types in Mandarin.

As Bai-Mandarin early bilingual children's L1-Bai uses neither pitch nor duration to encode focus types that differed in constituent size, we suggest that the unsuccessful acquisition of using prosodic cues for encoding focus types in the bilinguals' L2 might be explained by the negative L1 transfer. Further, non-Standard Mandarin-like input might also play an important role. Specifically, Bai-Mandarin early bilingual children grow up in a bilingual community where non-Standard varieties of Mandarin are spoken, including a learner variety of Mandarin, i.e., Semi-Standard Mandarin produced by Bai-Mandarin early bilingual adults (see Chapter 5), and a regional variety of Mandarin, i.e., Dali Mandarin (see Chapter 3). None of these

varieties of Mandarin mark focus in a Standard Mandarin-like manner (see Chapter 3 and 5). These non-Standard varieties of Mandarin compose the major input received by Bai-Mandarin early bilingual children in everyday communication (more on this in Chapter 6).

To sum up, this is the first investigation into the developmental trajectory of prosodic focus-marking in bilingual children's L2. We have found that Bai-Mandarin early bilingual children acquire the use of duration for focus-marking purposes earlier relative to the use of pitch-related cues, like monolingual Mandarin-speaking children. However, Bai-Mandarin early bilingual children have not acquired the use of prosody in distinguishing focus types earlier than in distinguishing narrow focus and non-focus, unlike monolingual Mandarin-speaking children. Both the L1 influence and non-Standard L2-input play important roles in early bilinguals' L2 development.[29]

[29] One should note, however, that there is a possible influence of language modes (i.e., only one language is active or both languages are active) on early bilingual children's linguistic performance (Genesee, 1989; Grosjean, 1997; Lanza, 1992; Treffers-Daller, 1997). In the current study, our experimenter is a native speaker of Standard Mandarin who only speaks Standard Mandarin with the participants. It is possible that the use of a Standard Mandarin experimenter in a school setting in the current study might have pushed the Bai-Mandarin early bilingual children closer to Standard Mandarin, reducing the difference with the monolinguals.

Table 25. Overview of the developmental trajectory of prosodic focus-marking Bai-Mandarin early bilingual children's Mandarin (present study) and monolingual Mandarin-speaking children's Mandarin (Yang, 2017; Yang & Chen, 2017)[30]

Age of Monolingual	Prosodic Cue	Focus vs. Pre-focus		Focus vs. Post-focus		Focus vs. Broad		Focus vs. Contrastive		Age of Bilingual
		Monolingual	Bilingual	Monolingual	Bilingual	Monolingual	Bilingual	Monolingual	Bilingual	
4-5	Duration	+	+	+	+	+	-	-	-	6-7
	Pitch span	-	-	-	-	-	-	-	-	
	Pitch-max	-	-	> (T1,T2,T4)	-	+	-	-	-	
	Pitch-min	< (T2)	-	> (T1)	-	-	-	-	-	
7-8	Duration	+	+	+	+	+	-	-	-	9-10
	Pitch span	+	-	> (T2,T4)	-	-	-	-	-	
	Pitch-max	> (T1,T4)	-	> (T1,T2,T4)	-	-	-	-	-	
	Pitch-min	< (T2,T3)	-	> (T1)	-	-	-	-	-	
10-11	Duration	+	+	+	+	+	-	+	-	12-13
	Pitch span	+	-	> (T2,T3,T4)	-	+	-	-	-	
	Pitch-max	+	-	> (T1,T2,T4)	-	-	-	-	-	
	Pitch-min	+	-	> (T1)	-	-	-	-	-	

[30] In Table 25, "+" indicates that the corresponding prosodic cue, such as duration, pitch span, pitch maximum or pitch minimum is used for distinguishing different focus conditions. "-" indicates that there is no evidence that the corresponding prosodic cue is used for differentiating different focus conditions. "T1", "T2", "T3", or "T4" indicates that the use of prosodic cues for distinguishing focus conditions is found in the indicated lexical tones; and "<" or ">" indicates the direction of the differences between different focus conditions.

CHAPTER 5 Ultimate Attainment in Bai-Mandarin Early Bilingual's Prosodic Focus-marking in Mandarin[31]

Abstract

This study investigates ultimate attainment in Bai-Mandarin early bilinguals' prosodic focus-marking in Mandarin. Mandarin SVO sentences with varied information structure were elicited through a picture-matching task. Participants were teachers of Standard Mandarin at a primary school in a Bai-speaking community, representative of the most proficient speakers of L2 Mandarin. The use of prosodic cues for encoding focus, including duration, pitch span, pitch maximum, and pitch minimum, was examined in different lexical tones. Our data shows that Bai-Mandarin early bilingual adults lengthened the duration of the focal constituent to distinguish narrow focus from non-focus (i.e., pre-focus and post-focus) in their Mandarin. In addition, they expanded the pitch span and raised the pitch maximum of the focal constituent to distinguish narrow focus from post-focus. However, the bilingual adults used neither duration nor pitch to distinguish narrow focus from focus types that differed in the size and contrastiveness of the focal constituent (i.e., broad focus and contrastive focus). These results thus suggest that Bai-Mandarin early bilingual adults encode focus prosodically in their L2 Mandarin by using duration and pitch-related cues, but their use of pitch-related prosodic cues to encode focus is less systematic in comparison to monolingual speakers of Standard Mandarin.

Keywords: early bilingualism, prosody, focus, ultimate attainment, Bai, Mandarin

[31] Preliminary results of a portion of the data from this chapter were presented at the 8th International Conference on Speech Prosody, and published in the proceedings (Liu, Chen, & Van de Velde, 2016a).

5.1 Introduction

An important function of prosody is to highlight new information in a sentence (also known as focus). Across languages that use prosody for focus-marking, similarities and differences have been found in the employment of prosodic cues (Burdin et al., 2015). For example, pitch and duration are used in Standard Mandarin (Xu, 1999; Yang, 2017), English (Cooper et al., 1985) and Dutch (Chen, 2009) for encoding focus, while duration is identified as the only prosodic cue for marking focus in Cantonese (Bauer et al., 2001; Wu & Xu, 2010), Tsat (Wang et al., 2012) and Bai (Chapter 2 of this thesis). The use of prosodic cues for encoding focus is also language-specific. For example, in both Dutch and Standard Mandarin, the pitch span of a focal constituent is expanded in comparison to its counterpart in non-focused position for encoding narrow focus. In Dutch, the expansion of the pitch span is solely realized by lowering the pitch-minimum without varying the pitch-maximum (Chen, 2009; Hanssen et al., 2008). However, the expansion of the pitch span is mainly realized by raising the pitch-maximum in Standard Mandarin (Xu, 1999; Yang, 2017). Such cross-linguistic differences in prosodic focus-marking can pose challenges to late bilinguals' acquisition of the language-specific manners of focus-marking (He et al., 2011; Nava & Zubizarreta, 2008; Rasier, Hiligsmann, Caspers, & Van Heuven, 2010). However, our understanding of the acquisition of prosodic focus-marking in early bilinguals is still limited, as only a few studies have examined early bilinguals and provided different conclusions (Chen et al., 2014; Gut & Pillai, 2014; Wang et al., 2012). The present study aims to establish early bilinguals' ultimate attainment of prosodic focus-marking in their L2 by examining the case of Bai-Mandarin early bilinguals' Mandarin.

In the literature of bilingualism and language acquisition, the age of bilinguals at the time of acquisition results in the differentiation between "early bilingualism" and "late bilingualism" (Hoffmann, 2014, p. 18). The cut-off age of early bilingualism (i.e., a time point when the second language kicks in) has been under discussion for decades, ranging from three years old to puberty (Bialystok & Miller, 1999;

DeKeyser, 2000; Goodluck, 1986; Guasti, 2004; Johnson & Newport, 1989; Krashen, 1973; Lakshmanan, 1995; Lenneberg, 1967; Long, 1990; McLaughlin, 2012; Penfield & Roberts, 2014; Schwartz, 2004; Unsworth, 2005). In the present study, following the age criteria set in Unsworth (2005), early bilinguals are denoted as learners whose acquisition of the second language (hereafter L2) begins between four to seven years.[32] Late bilinguals are denoted as learners whose acquisition of the second language begins at an age of eight years or older.

5.1.1 Prosodic focus-marking in late bilinguals

Previous studies suggest that the use of prosodic cues for encoding focus in L2 is quite difficult to acquire for late bilinguals. On the one hand, differences in learners' L1 and L2 prosodic systems can hinder the acquisition of prosodic focus-marking (Backman, 1979; Gut & Pillai, 2014; Kelm, 1987; Nava & Zubizarreta, 2008, 2009, Rasier & Hiligsmann, 2007, 2009; Swerts & Zerbian, 2010; Turco et al., 2015; Ueyama & Jun, 1996; Zubizarreta & Nava, 2011). For example, in English, pitch accents are assigned for pragmatic purposes (e.g., encoding focus), whereas the placement of pitch accent in Spanish is determined by sentence structure (e.g., nuclear accent is assigned to the right-edge of the sentence) and seldom influenced by pragmatic purposes. Nava and Zubizarreta (2008) investigated prosodic focus-marking in English produced by Spanish learners and found that Spanish learners of English were unable to use accent placement to distinguish narrow focus from non-focus in their L2-English, especially for learners with low L2 proficiency. Swerts and Zerbian (2010) also found that only proficient Zulu (L1) learners of South African English (L2) were able to use prosody to encode focus, which was absent in their L1. On the other hand, similarities in the manner of using prosodic cues for encoding focus between learners' L1 and L2 do not necessarily lead to the successful learning of prosodic focus-marking in learners' L2 (Chen, 2014; He et al., 2011; McGory, 1997). For example, it is known that prosodic focus-marking in

[32] "Early bilinguals" defined in the present study is an equivalent term to "child L2 learners" defined in Paradis (2007). Specifically, child L2 learners are denoted as children "who have established one language before they begin learning the other, and typically speak the L1 at home and the L2 at school" (Paradis, 2007, p. 387).

Dutch and Standard Mandarin is similar from a phonetic perspective; that is, expanded pitch span and lengthened duration are used for encoding narrow focus. He et al. (2011) found that Mandarin learners of Dutch produced falling pitch patterns which were similar in shape to the falls produced by monolingual speakers of Dutch for signaling focus in their L2-Dutch. However, the phonetic realization of the falling pitch accents produced by Mandarin learners of Dutch regarding onset duration, rime duration, coda duration, nuclear fall pitch excursion, peak delay (i.e., the timing of the peak relative to the beginning of the vowel), and fall duration.

Although the difficulty of acquiring prosodic focus-marking is widely observed in late bilinguals, simultaneous bilinguals (i.e., bilinguals who grow up with two languages simultaneously from birth) are successful in exploiting prosodic cues for encoding focus in both of their languages (Wu & Chung, 2011), and heritage speakers (i.e., speakers who are raised in a home where a non-dominant language is spoken, who speak or merely understand the heritage language, and who are to some degree bilingual in dominant and the heritage language) are successful in exploiting prosodic cues for encoding focus in their L2s (Chen, 2014; Hoot, 2012). The successful acquisition of prosodic focus-marking by simultaneous bilinguals and heritage speakers has been attributed to their early exposure to the target language (Chen, 2014; Wu & Chung, 2011).

5.1.2 Prosodic focus-marking in early bilinguals

Only a few studies have examined early bilinguals and provided different conclusions. For example, Huang and Jun (2011) found that Mandarin learners of English (Age of acquisition or AoA: from five to seventeen years old) were largely native-like in their use of accent placement to encode focus prosodically in their L2. Chen (2014) and Chen et al. (2012, 2014) also found that young Quanzhou Southern Min-Mandarin (hereafter, QZSM-Mandarin) bilinguals who received more Standard Mandarin input and intensive training in Standard Mandarin produced the post-focal constituent with compressed pitch span, like monolingual speakers of Standard Mandarin. However, other studies showed that the acquisition of prosodic focus-

marking in early bilinguals was unsuccessful. For example, Wang et al. (2012) found that Tsat-Mandarin early bilinguals did not vary pitch to encode focus in their Mandarin (L2), unlike monolingual speakers of Standard Mandarin. Thus, Tsat-Mandarin early bilinguals did not show Mandarin-native-like ability to encode focus prosodically, although they were comparable to QZSM–Mandarin bilinguals examined in Chen (2014) in terms of living in a target-language-dominant environment (i.e., the Mandarin-dominant environment of China). Also, Gut and Pillai (2014) showed that, unlike monolingual speakers of English, Malay (L1) learners of English (L2) only showed durational variation for encoding focus in their English, despite that they started their L2-English learning in childhood (AoA from four to seven years). These mixed findings might be due to the differences in L2-input and use between the early bilinguals examined in different studies. Specifically, Chen et al. (2014) showed that only the young QZSM-Mandarin bilinguals who received more Standard Mandarin-like input of their L2-Mandarin education than the old QZSM -Mandarin bilinguals who received less Standard Mandarin-like input in their L2-Mandarin education varied prosodic cues for encoding focus in a Standard Mandarin-like manner in their L2-Mandarin. However, the Tsat-Mandarin bilinguals were brought up and had been living in a Tsat village and Malay-English early bilinguals had been living in Malay where English (L2) was primarily used for academic purposes. The importance of the characteristics of L2-input has also been found in early bilinguals' L2 acquisition in other linguistic domains (Flege, 2009; Flege & Liu, 2001; Piske et al., 2001). Together, these studies suggest that early exposure to L2, extensive use of L2, and quality of L2 training may be crucial to whether early bilinguals can obtain native-like competence in L2, in line with findings on the acquisition of vocabulary and syntax in early bilinguals (Dulay & Burt, 1973; Jia, 2003; Kohnert, 2004; Paradis, 2005).

On a methodological note, the above-mentioned studies are limited in several ways. First, the effect of focus type was not investigated. It is known that focus types that differed in the size and contrastiveness of the focal constituent are prosodically differentiated in many languages (Chen & Gussenhoven, 2008; De Jong, 1980;

Gussenhoven, 1983, 2007, 2008; Ouyang & Kaiser, 2015). Second, recent studies of monolingual Mandarin-speaking children's L1 acquisition have shown that lexical tones interact with the acquisition of prosodic focus-marking. That is, the use of prosodic cues for encoding focus in some lexical tones are acquired earlier than in other lexical tones for Standard Mandarin-speaking monolingual children (Yang & Chen, 2017, 2014). Given that lexical tones were not systematically examined in previous studies, it is difficult to know whether earlier findings on the use of prosody to mark focus in certain tones are generalizable to all lexical tones in the target language. Finally, it has been shown that the prosodic realization of focus differs in read speech and naturalistic speech (Bard & Aylett, 1999; Chen & Gussenhoven, 2008; De Ruiter, 2010; O'Brien & Gut, 2010; Xu, 1999; Yang, 2017). For example, Xu (1999) showed that the duration and pitch span of the focal constituent is lengthened and expanded for encoding narrow focus in read speech in Standard Mandarin, as found in semi-spontaneous speech in Standard Mandarin (Yang, 2017). However, Chen and Gussenhoven (2008) showed that both pitch and duration are exploited to distinguish contrastive focus from narrow focus in read speech in Standard Mandarin, while Yang (2017) found that neither pitch nor duration is used for the same purpose in semi-spontaneous speech in Standard Mandarin. The question arises as to how successful early bilinguals are in realising focus prosodically in semi-spontaneous speech in their L2-Mandarin.

5.1.3 The present study

In the present study, we examine prosodic focus-marking in Mandarin by a group of Bai-Mandarin early bilingual adults who are representative of the most proficient speakers of Standard Mandarin among Bai-Mandarin early bilinguals (i.e., teachers of Standard Mandarin at a primary school). These early bilinguals formally have acquired Standard Mandarin at an early age (six to seven years old). Informally, they have exposed to Dali Mandarin (a regional variety of Mandarin) in their daily life before the age of six. Further, they had intensive training in Standard Mandarin, use Standard Mandarin actively in their daily lives, and have high Standard Mandarin proficiency.

Bai is Bai-Mandarin early bilinguals' L1. As a Tibetan-Burman tone language, Bai is spoken by more than one million Bai people, mainly in Dali Bai Autonomous Prefecture, Yunnan Province, China. As mentioned earlier, focus is prosodically encoded via pitch and duration in Standard Mandarin (Chen, 2010; Chen & Gussenhoven, 2008; Ouyang & Kaiser, 2015; Xu, 1999; Yang, 2017; Yang & Chen, 2014). Regarding the use of prosodic cues for encoding narrow focus, the pitch span and duration of the focal constituent are expanded and lengthened relative to the non-focal counterpart (Yang, 2017; Yang & Chen, 2014). Further, the pitch span of the post-focal constituent is compressed relative to the non-focal counterpart in broad focus condition (Xu, 1999). Regarding the use of prosodic cues for encoding focus types that differed in the size of the focal constituent, the pitch span and duration of the focal constituent are expanded and lengthened in narrow focus, relative to its counterpart in broad focus in read speech (Xu, 1999). However, in semi-spontaneous speech, only the duration of the focal constituent is lengthened in narrow focus to distinguish from broad focus, while the pitch span is not varied (Yang, 2017; Yang & Chen, 2014). Regarding the use of prosodic cues for encoding focus types that differed in the contrastiveness of the focal constituent, the pitch span and duration of the focal constituent are expanded and lengthened in contrastive focus, relative to its counterpart in non-focus in read speech (Chen & Gussenhoven, 2008). However, in semi-spontaneous speech, neither pitch span nor duration is used to distinguish contrastive focus from narrow focus in semi-spontaneous speech (Yang, 2017; Yang & Chen, 2017). Unlike Standard Mandarin, only duration is used as a prosodic cue for encoding focus in Bai (Chapter 2). Specifically, the duration of the focal constituent is lengthened to distinguish narrow focus from post-focus, while pitch-related prosodic cues are not varied for encoding narrow focus in Bai. Further, neither pitch-related cues nor duration is varied to encode focus types in Bai.

We aim to address the question of whether Bai-Mandarin early bilinguals can ultimately attain native-like ability in using prosody for encoding focus in their L2 Mandarin. Given that early exposure to L2, extensive use of L2 and high quality of

L2 training appear to determine whether early bilinguals can obtain native-like competence in L2 (Chen et al., 2014; Gut & Pillai, 2014; Huang & Jun, 2011; Wang et al., 2012), we hypothesize that the Bai-Mandarin early bilingual adults under investigation should have native-like competence in prosodic focus-marking in Mandarin. Our predictions are that the Bai-Mandarin early bilingual adults are native-like in the use of duration and pitch-related cues to distinguish narrow focus from non-focus in all lexical tones, and in the use of duration to distinguish narrow focus from broad focus in all lexical tones in L2 Mandarin.

5.2 Method

5.2.1 Picture-matching game

We adopted the picture-matching game used in Yang and Chen (Yang & Chen, 2017, 2014) and Yang (2017) to elicit SVO sentences in Mandarin from Bai-Mandarin early bilingual adults in an interactional setting. Three sets of pictures were used. The experimenter and the participant each held a set of pictures ordered in a specific sequence; and the third set of pictures was scattered around on a table. In the experimenter's pictures (the first set), there was always some information missing, e.g., a subject, an action (verb), an object, or all three. The participant's pictures (the second set) all contained a complete event. The participant's task was to help the experimenter with sorting out pictures from the first set (experimenter's pictures) and the third set (scattered around on the table) in matched pairs (see Figure 1).

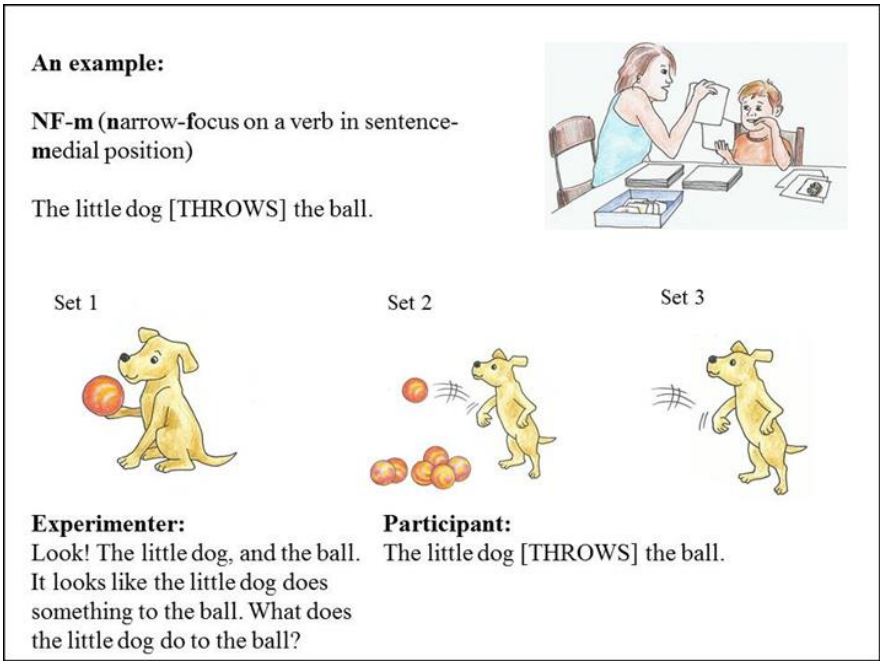

Figure 1. An example of a trial eliciting the sentence "The little dog [THROWS] the ball".

An example of a trial eliciting a target sentence with **n**arrow-**f**ocus on a verb in sentence-**m**edial position (the NF-m condition): First, the experimenter took a picture (e.g., a dog and a ball) from her own set, drew the participant's attention to the picture and established what the picture was by saying, e.g., "Look! The dog and the ball. It looks like the dog does something to the ball." This was done to make sure that the entity in the picture was referentially given to the participant before the utterance of the question. Second, the experimenter asked a question about the picture (e.g., "What does the dog do to the ball?"). Third, the participant took a picture from his or her set and inspected it. The experimenter then repeated the question, followed by an answer from the participant (e.g., "The dog [THROWS] the ball."). Lastly, the experimenter found the picture containing the missing information in the third set and paired it up with her own picture. As to rules of the

game, participants were explicitly asked to respond in full sentences and not to show their own pictures to the experimenter.

To ensure consistency in the participants' word choice, the picture-matching game was preceded by a picture-naming task, which was designed to familiarize the participants with the target words and the entities in the pictures used in the game.

5.2.2 Experimental materials

SVO Mandarin sentences in five focus conditions were elicited via the picture-matching game: **n**arrow-**f**ocus on the subject NP in sentence-**i**nitial position (NF-i); **n**arrow-**f**ocus on the verb in sentence-**m**edial position (NF-m); **n**arrow-focus on the object NP in sentence-**f**inal position (NF-f); **b**road **f**ocus (BF) and **c**ontrastive-**f**ocus on the verb in sentence-**m**edial position (CF-m). The focus condition was set up by a WH-question or a statement from the experimenter, as illustrated in examples (1) to (5).

(1) **NF-i**

Experimenter: 看！球。球飞在空中。看起来有小动物扔球。谁扔球？

Look! The ball. The ball is in the air. It looks like someone throws the ball. Who throws the ball?

Participant: [小熊]扔球。

[THE LITTLE BEAR] throws the ball.

(2) **NF-m**

Experimenter: 看！小狗，还有球。看起来小狗要弄球。小狗怎么弄球？

Look! The little dog, and the ball. It looks like the little dog does something to the ball. What does the little dog do to the ball?

Participant: 小狗[扔]球。

The little dog [THROWS] the ball.

(3) **NF-f**

Experimenter: 看！小兔。小兔的胳膊挥出去了。看起来小兔在扔东西。小兔扔什么？

Look! The little rabbit, and it stretches out its arm. It looks like the little rabbit throws something. What does the little rabbit throw?

Participant: 小兔扔[球] 。

The little rabbit throws [THE BALL].

(4) **BF**

Experimenter: 看！阿姨什么都看不清。你的图片上讲了什么？

Look! My picture is very blurry. I cannot see anything clearly. What happens in your picture?

Participant: [小熊扔球] 。

[THE LITTLE BEAR THROWS THE BALL].

(5) **CF-m**

Experimenter: 看！小猫，还有球。看起来小猫要弄球。我猜，小猫剪球。

Look! The little cat, and the ball. It looks like the little cat will do something to the ball. I will make a guess: The little cat [CUTS] the ball.

Participant: 小猫[扔] 球。

The little cat [THROWS] the ball.

The SVO sentences were constructed in such a way that each was a unique combination of a subject-noun and a VP (verb + object-noun). Each of the four

lexical tones (i.e., high-level tone/Tone 1, rising tone/Tone 2, dipping tone/Tone 3, and falling tone/Tone 4) occurred in subject noun phrases, verbs, and object nouns. Four monosyllabic verbs and four monosyllabic object nouns formed 16 VPs, each of which appeared in each focus condition (n = 5). This resulted in 80 VPs, and then four subject-noun phrases, which started with the word "xiǎo (little) and ended with four nouns carrying four lexical tones, were evenly distributed over these 80 VPs to form 80 SVO target sentences. An overview of the words in SVO sentences is provided in Table 1.

Table 1. Overview of the words in SVO sentences, each word is presented in Chinese, Pinyin and English translation

	Subject	Verb	Object
Tone 1	小猫 xiǎo māo cat (little cat)	扔 rēng throw	书 shū book
Tone 2	小熊 xiǎo xióng bear (little bear)	埋 mái bury	球 qiú ball
Tone 3	小狗 xiǎo gǒu dog (little dog)	剪 jiǎn cut	笔 bǐ pen
Tone 4	小兔 xiǎo tù rabbit (little rabbit)	运 yùn transport	菜 cài vegetable

5.2.3 Participants and procedure

Six Bai-Mandarin early bilinguals (four women and two men, aged between 29 and 51 years, average age = 43, SD = 7.4) participated in our experiment (see Appendix D). All the bilingual participants acquired Bai at home from birth and started their formal Standard Mandarin education at the age of six.[33] Informally, they have been

[33] We observed that Bai, rather than Mandarin was used in the preschool in Xizhou County. This observation was confirmed through an interview with the school principle of Jinhe Preschool who was also the school principle of Jinhe Primary School. Our participants either

exposed to Dali Mandarin (a regional variety of Mandarin) in their daily life before the age of six. Also, all of them were teachers of Standard Mandarin at a primary school in a local Bai-speaking community (Jinhe Primary School, Xizhou County, Dali Bai Autonomous Prefecture, China) at the time of testing. They had been living in the Bai community for their lifetime until the time of testing. All of them either graduated from a teachers' college or held certificates for teaching Standard Mandarin and for advanced Standard Mandarin proficiency. They had received intensive training in Standard Mandarin in order to meet the requirements of their professions. In addition, they had been actively using Standard Mandarin in their daily lives.

The participants were tested individually by a female experimenter (aged 27 years), who was a native speaker of Standard Mandarin and only spoke Standard Mandarin with the participants. Each experiment was split into two sessions, and each session contained 40 trials. Each experiment took about 25-30 minutes for a participant to complete. All the test sessions were conducted in a quiet room in Jinhe Primary School, Xizhou County. The experiments were recorded using a portable ZOOM H1 digital recorder at a 44.1 kHz sampling rate and 16-bit accuracy. Each session was also video-recorded for future training purposes.

5.2.4 Acoustic annotation

The auditory recordings from the participants were first orthographically annotated. An answer sentence was discarded for analysis if: (1) it was not uttered as a direct response to the target question; (2) it contained no self-corrections or hesitations (defined as a long "em" sound produced by the participants before answering questions); (3) it deviated from the target sentences in word choice or word order. In total, 90% of the obtained responses (n = 432) were included for further analysis.

---

attended the preschool where Bai was used or had not attended the preschool at all. All of the participants claimed that they started their formal Mandarin education from the first year of their primary school (i.e., at the age of six years).

The **verbs** were the targets of prosodic analysis. They were focal constituents in the BF, NF-m, and CF-m conditions, pre-focal constituents in the NF-f condition, and post-focal constituents in the NF-i condition. This setup made it possible to investigate effects of both focus and focus types on pitch and duration in the verbs. The verbs were acoustically annotated by examining the waveform, wide-band spectrum, and pitch track in Praat (Boersma & Weenink, 2006) in combination with auditory impressions (Turk et al., 2006). Two pitch-related landmarks and two segmental landmarks were labelled in each verb: pitch maximum, pitch minimum, word onset, and word offset.

The tonal targets of the lexical tones in Standard Mandarin were taken into consideration when the landmarks for pitch maximum and pitch minimum were inserted (Xu, 1997; Xu & Wang, 2001; Yang & Chen, 2014). Specifically, Tone 1 is a high-level tone, but we observed that it appeared as a slightly rising pitch contour and occasionally as a slightly falling pitch contour sentence-medially. The pitch-maximum was labelled after the pitch-minimum in the case of a rising Tone 1 syllable but before the pitch-minimum in the case of a falling Tone 1 syllable. Tone 2 is a rising tone, but we observed that it appeared as a fall-rise contour sentence-medially, as reported in Xu (1997) and Yang and Chen (2014). As the falling part is largely influenced by the preceding tone, the rising part is taken to contain the tonal targets of Tone 2 (Xu, 1997). Thus, the pitch-maximum was labelled after the pitch-minimum in the rising part of Tone 2. Similarly, Tone 4 is a falling tone, but we observed that it appeared as a rise-fall contour sentence-medially, again, as reported in Xu (1997) and Yang and Chen (2014). As the rising part is subjected to the influence of the preceding tone, the falling part contains the tonal targets of Tone 4. Thus, the pitch-maximum was labelled before the pitch-minimum in the falling part of Tone 4. Tone 3 is a fall-rise tone when produced in isolation or sentence-finally, but it was realized as a falling pitch contour sentence-medially, and as a fall-rise contour when followed by another Tone 3, as reported in Xu (1997) and Yang and Chen (2014). We labelled the pitch-maximum before the pitch-minimum in the case

of falling Tone 3 but after the pitch-minimum in the rising part of a falling-rising Tone 3.

The pitch values (in Hz) at the pitch-related landmarks and time values (in seconds) at the segmental landmarks were automatically extracted by means of Praat scripts. Four measures were obtained from each verb: pitch maximum, pitch minimum, pitch span (i.e., the difference between the maximum pitch and the minimum pitch), and word duration (i.e., offset time minus onset time). In 40 of the usable responses (9%), an accurate measurement of pitch values was not possible. These responses were thus excluded from the analysis of pitch-related measurements.

5.3 Statistical analyses and results

5.3.1 Statistical analyses

In order to assess the effect of focus, we compared the measurements of the verbs between the unfocused conditions and the focused condition; that is, NF-m (verb is the focal constituent) vs. NF-i (verb is the post-focal constituent), and NF-m (verb is the focal constituent) vs. NF-f (verb is the pre-focal constituent). To investigate the effect of focus type that differed in the size of the focal constituent, we compared narrow focus on the verb (NF-m) with broad focus (BF). To investigate the effect of focus type that differed in the contrastiveness of the focal constituent, we compared contrastive focus on the verb (CF-m) with non-contrastive focus on the verb (NF-m).

Statistical analyses were conducted using mixed-effects modelling in R (R Core Team, 2014) with package "lme4" (Bates, Mächler, et al., 2015) and "lmerTest" (Kuznetsova et al. 2013). In all models, tone and focus conditions were included as fixed factors, while speaker (i.e., the participants) and sentence (i.e., the answer sentences) were included as random factors. In each of the four pair-wise comparisons listed above, focus condition had two levels (the two focus conditions of interest), and tone referred to the lexical tones of the verbs, which had four levels (Tones 1, 2, 3, and 4). Outcome variables were the duration, pitch span, pitch maximum, and pitch minimum of the verbs.

Following Magezi (2015) and Field et al. (2012), our models were constructed and evaluated in a stepwise fashion. Specifically, we started from an intercept-only model, built the subsequent models by adding one new term/factor at a time, and then systematically compared the models which differed only by the presence or absence of one term. The comparison was done using a likelihood-ratio test, and the test statistic χ^2, degrees of freedom and *p*-value were reported for the added term. A *p*-value of less than 0.05 was considered to indicate that the added term attributed significantly to the model fit (Magezi, 2015). A detailed description of the model building is presented in Table 2.

Table 2. Model build-up procedure

Model	Added term
Model 0	only contains speaker, sentence as random intercepts
Model 1	+ tone
Model 2	+ focus condition
Model 3	+ tone : focus condition

When building the models, only the factors and interactions that significantly improved the fit of the model were retained until the best-fit model was determined. After the best-fit model was established, we only summarized and interpreted the model with the best fit. For each analysis, we reported the results of the model comparison first, and then the parameter estimates of the best-fit model. As we are primarily interested in the effect of focus on the outcome variables, we will concentrate on the main effect of focus or interactions involving the factor of focus. However, we do not discuss in detail the main effects of the factors when the interactions involving these factors are significant.

5.3.2 Duration

5.3.2.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The effect of focus (focus vs. post-focus) on duration was tested with the models specified in Table 2. The results of the fit and model comparisons are shown in Table 3. As can be seen, the best-fit model was Model 2, which contained the main effects of tone, $\chi^2(3) = 13.366$, $p < .01$, and focus condition, $\chi^2(1) = 12.636$, $p < .001$. Parameter estimates of the best-fit model are presented in Table 4. The main effect of the focus condition was such that the duration of the verbs was significantly longer in narrow focus (259.8 ms, SD = 54.7) than in post-focus (237.7 ms, SD = 47.5), regardless of tone ($b = 20.604$, $df = 30.94$, $t = 3.892$, $p < .001$), as shown in Figure 2.

Table 3. Duration, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−912.80				
1 + tone	7	−906.12	0 vs 1	13.366	3	.014**
2 + focus condition	8	−899.80	1 vs 2	12.636	1	.000***
3 + tone : focus condition	11	−899.21	2 vs 3	1.167	3	.761

Table 4. Duration, narrow focus (NF-m) vs. post-focus (NF-i), parameter estimates of the best-fit model (Model 2)

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	238.575	15.469	7.603	15.432	.000***
Tone 2	19.150	15.469	7.603	15.423	.015*
Tone 3	−18.298	7.422	29.996	−2.466	.020*
Tone 4	0.769	7.514	31.430	0.102	.919
Narrow focus	20.604	5.293	30.944	3.892	.000***
Random part	*name*	S^2	*SE*		
Sentence	Intercept	7.89	2.809		
Speaker	Intercept	1230.36	35.076		
Residual		1204.72	34.709		

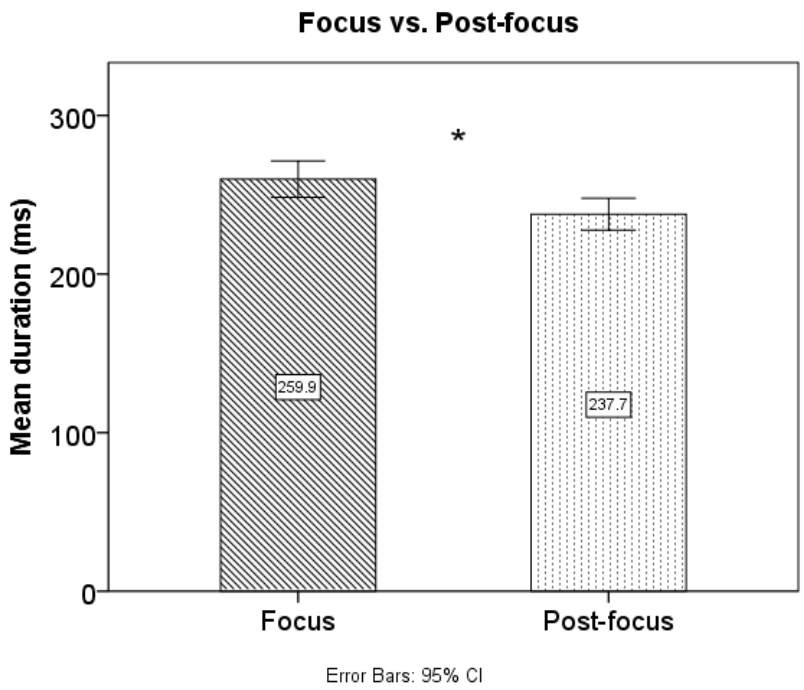

Figure 2. Mean duration (in ms) of focal constituent vs. post-focal constituent, n =179, N = 6. Significant differences are marked with an asterisk.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 5. As can be seen, the best-fit model was Model 2, which contained the main effects of tone, χ^2 (3) = 13.802, p < .01, and focus condition, χ^2 (1) = 10.319, p < .01. Parameter estimates of the best-fit model are presented in Table 6. The main effect of focus condition was such that the duration of the verbs was significantly longer in narrow focus (259.9 ms, SD = 54.7) than in pre-focus (238.1 ms, SD = 49), regardless of tone (b = 22.191, df = 30.69, t = 3.471, p < .01), as shown in Figure 3.

Table 5. Duration, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−931.92				
1 + tone	7	−925.01	0 vs 1	13.802	3	.003**
2 + focus condition	8	−919.85	1 vs 2	10.319	1	.001**
3 + tone : focus condition	11	−918.44	2 vs 3	2.820	3	.420

Table 6. Duration, narrow focus (NF-m) vs. pre-focus (NF-f), parameter estimates of the best-fit model (Model 2)

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	241.167	15.825	8.764	15.240	.000***
Tone 2	16.955	9.009	30.322	1.882	.069
Tone 3	−27.466	9.040	30.630	−3.038	.005**
Tone 4	−2.862	9.007	30.299	−0.318	.753
Narrow focus	22.191	6.392	30.694	3.471	.002**
Random part	*name*	S^2	*SE*		
Sentence	Intercept	7.89	2.809		
Speaker	Intercept	1230.36	35.076		
Residual		1204.72	34.709		

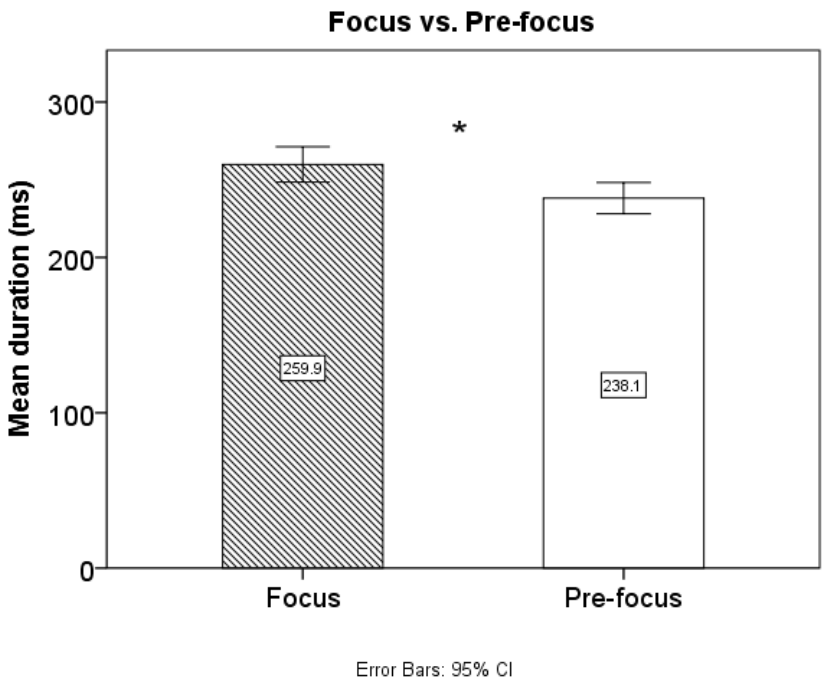

Figure 3. Mean duration (in ms) of focal constituent vs. pre-focal constituent, n = 184, N = 6. Significant differences are marked with an asterisk.

5.3.2.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 7. As can be seen, the best-fit model was Model 1, which contained the main effect of tone, χ^2 (3) = 16.606, $p < .001$. There was thus no evidence that the Bai-Mandarin early bilingual adults varied duration in their Mandarin to distinguish narrow focus from broad focus, regardless of tone.

Table 7. Duration, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−878.38				
1 + tone	7	−870.08	0 vs 1	16.606	3	.000***
2 + focus condition	8	−868.95	1 vs 2	2.247	1	.134
3 + tone : focus condition	11	−868.19	2 vs 3	1.533	2	.675

5.3.2.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 8. As can be seen, the best-fit model was Model 1, which contained the main effect of tone, χ^2 (3) = 14.148, $p < .01$. There was thus no evidence that the Bai-Mandarin bilingual adults varied duration in their Mandarin to distinguish contrastive focus from narrow focus (i.e., non-contrastive focus), regardless of tone.

Table 8. Duration, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−855.23				
1 + tone	7	−848.15	0 vs 1	14.148	3	.003**
2 + focus condition	8	−847.77	1 vs 2	0.765	1	.382
3 + tone : focus condition	11	−847.33	2 vs 3	0.890	3	.828

5.3.3 Pitch span

5.3.3.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 9. As can be seen, the best-fit model was Model 2, which contained the main effects of tone, χ^2 (3) = 19.805, $p < .001$, and focus condition, χ^2 (1) = 4.338, $p < .05$. Parameter estimates of the best-fit model are presented in Table 10. The main effect of the focus condition was such that the pitch span of the verbs was significantly wider in narrow focus (40.2 Hz, SD = 19.4) than in post-focus (32.7 Hz, SD = 18.3), regardless of tone (b = 7.701, df = 21.01, t = 2.429, $p < .05$), as shown in Figure 4.

Table 9. Pitch span, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−688.58				
1 + tone	7	−678.68	0 vs 1	19.805	3	.000***
2 + focus condition	8	−676.51	1 vs 2	4.338	1	.037*
3 + tone : focus condition	11	−675.85	2 vs 3	1.329	3	.722

Table 10. Pitch span, narrow focus (NF-m) vs. post-focus (NF-i), parameter estimates of the best-fit model (Model 2).

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	25.316	3.730	27.940	6.788	.000***
Tone 2	11.420	4.449	30.900	2.567	.015*
Tone 3	4.948	4.466	31.230	1.108	.276
Tone 4	23.961	4.394	29.560	5.453	.000***
Narrow focus	6.833	3.170	31.630	2.155	.039*
Male	−7.417	3.187	6.590	−2.327	.055.
Random part	*name*	S^2	*SE*		
Sentence	Intercept	38.072	6.170		
Speaker	Intercept	5.251	2.292		
Residual		209.449	14.472		

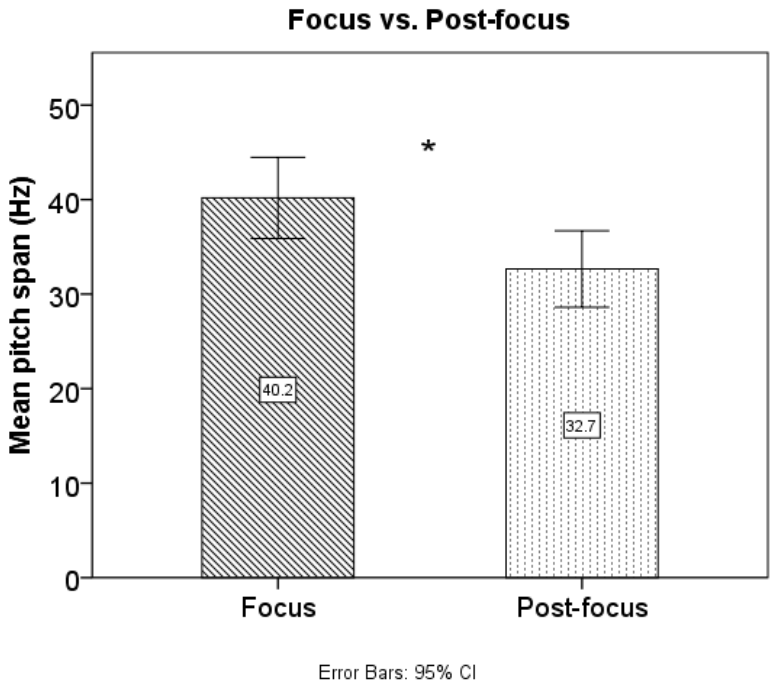

Figure 4. Mean pitch span (in Hz) of focal constituent vs. post-focal constituent, n = 162, N = 6. Significant differences are marked with an asterisk.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 11. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (3) = 23.026, p < .001. There was thus no evidence that Bai-Mandarin bilingual adults varied pitch span to distinguish narrow focus from pre-focus in their Mandarin, regardless of tone.

Table 11. Pitch span, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−694.42				
1 + tone	7	−682.91	0 vs 1	23.026	3	.000***
2 + focus condition	8	−682.70	1 vs 2	0.424	1	.515
3 + tone : focus condition	11	−682.15	2 vs 3	1.102	3	.777

5.3.3.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 12. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, $\chi^2(3) = 28.912$, $p < .001$. There was thus no evidence that the Bai-Mandarin bilingual adults varied pitch span to distinguish narrow focus from broad focus in their Mandarin, regardless of tone.

Table 12. Pitch span, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−678.73				
1 + tone	7	−664.27	0 vs 1	28.912	3	.000***
2 + focus condition	8	−663.76	1 vs 2	1.024	1	.312
3 + tone : focus condition	11	−663.22	2 vs 3	1.079	3	.782

5.3.3.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 13. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (3) = 22.009, p < .001. There was thus no evidence that the Bai-Mandarin bilingual adults varied pitch span to distinguish contrastive focus from narrow focus (i.e., non-contrastive focus) in their Mandarin, regardless of tone.

Table 13. Pitch span, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−666.33				
1 + tone	7	−655.33	0 vs 1	22.009	3	.000***
2 + focus condition	8	−655.03	1 vs 2	0.587	1	.444
3 + tone : focus condition	11	−654.42	2 vs 3	1.221	3	.748

5.3.4 Pitch maximum

5.3.4.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 14. As can be seen, the best-fit model was Model 2, which contained the main effects of tone, χ^2 (3) = 68.973, p < .001, and focus condition, χ^2 (1) = 5.553, p < .05. Parameter estimates of the best-fit model are presented in Table 15. The main effect of focus condition was such that the pitch maximum of the verbs was significantly higher in narrow focus (198.4 Hz, SD = 59.2) than in post-focus (191.4 Hz, SD = 57.6), regardless of tone (b = 6.287, df = 31.99, t = 2.56, p < .05), as shown in Figure 5.

Table 14. Pitch maximum, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−740.06				
1 + tone	7	−705.57	0 vs 1	68.973	3	.000***
2 + focus condition	8	−702.79	1 vs 2	5.553	1	.018*
3 + tone : focus condition	11	−702.73	2 vs 3	0.127	3	.988

Table 15. Pitch maximum, narrow focus (NF-m) vs. post-focus (NF-i), parameter estimates of the best-fit model (Model 2)

	Estimate	SE	df	t value	Pr (>|t|)
Fixed part					
Intercept	207.27	22.017	6.150	9.414	.000***
Tone 2	−45.798	3.682	156.01	−12.439	.000***
Tone 3	−35.759	3.698	156.00	−9.670	.000***
Tone 4	5.029	3.597	156.01	1.398	.164
Narrow focus	6.359	2.631	156.00	2.417	.017*
Random part	*name*	S^2	*SE*		
Sentence	Intercept	0.0	0.00		
Speaker	Intercept	2861.7	53.49		
Residual		278.8	16.70		

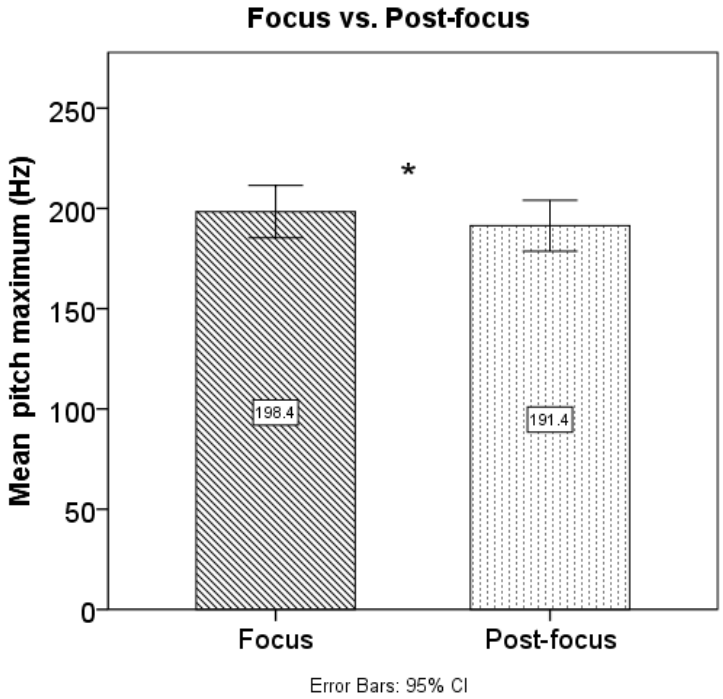

Figure 5. Mean pitch maximum (in Hz) of focal constituent vs. post-focal constituent, n = 162, N = 6. Significant differences are marked with an asterisk.

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 16. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (3) = 52.607, p < .001. There was thus no evidence that the Bai-Mandarin early bilingual adults varied pitch maximum to distinguish narrow focus from pre-focus in their Mandarin, regardless of tone.

Table 16. Pitch maximum, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−735.93				
1 + tone	7	−709.63	0 vs 1	52.607	3	.000***
2 + focus condition	8	−709.58	1 vs 2	0.097	1	.756
3 + tone : focus condition	11	−709.43	2 vs 3	0.306	3	.959

5.3.4.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 17. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, $\chi^2(3) = 68.385$, $p < .001$. There was thus no evidence that the Bai-Mandarin early bilingual adults varied pitch maximum to distinguish narrow focus from broad focus in their Mandarin, regardless of tone.

Table 17. Pitch maximum, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−724.12				
1 + tone	7	−689.93	0 vs 1	68.385	3	.000***
2 + focus condition	8	−689.75	1 vs 2	0.365	1	.546
3 + tone : focus condition	11	−689.05	2 vs 3	1.397	3	.706

5.3.4.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 18. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (3) = 61.876, p < .001. There was thus no evidence that the Bai-Mandarin bilingual adults varied pitch maximum to distinguish contrastive focus from narrow focus (i.e., non-contrastive focus) in their Mandarin, regardless of tone.

Table 18. Pitch maximum, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−710.65				
1 + tone	7	−679.72	0 vs 1	61.876	3	.000***
2 + focus condition	8	−679.63	1 vs 2	0.184	1	.663
3 + tone : focus condition	11	−677.86	2 vs 3	3.521	3	.318

5.3.5 Pitch minimum

5.3.5.1 Effect of focus: focus vs. non-focus

Narrow focus vs. post-focus (NF-m vs. NF-i). The results of the fit and model comparisons are shown in Table 19. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (3) = 50.165, p < .001. There was thus no evidence that the Bai-Mandarin bilingual adults varied pitch minimum to distinguish narrow focus from post-focus in their Mandarin, regardless of tone.

Table 19. Pitch minimum, narrow focus (NF-m) vs. post-focus (NF-i), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−732.46				
1 + tone	7	−707.38	0 vs 1	50.165	3	.000***
2 + focus condition	8	−705.88	1 vs 2	5	1	.083.
3 + tone : focus condition	11	−705.16	2 vs 3	1.443	3	.695

Narrow focus vs. pre-focus (NF-m vs. NF-f). The results of the fit and model comparisons are shown in Table 20. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (3) = 50.697, $p < .001$. There was thus no evidence that the Bai-Mandarin bilingual adults varied pitch minimum to distinguish narrow focus from pre-focus in their Mandarin, regardless of tone.

Table 20. Pitch minimum, narrow focus (NF-m) vs. pre-focus (NF-f), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−731.23				
1 + tone	7	−705.88	0 vs 1	50.697	3	.000***
2 + focus condition	8	−705.80	1 vs 2	0.161	1	.688
3 + tone : focus condition	11	−705.31	2 vs 3	0.979	3	.806

5.3.5.2 Effect of size: narrow focus (NF-m) vs. broad focus (BF)

The results of the fit and model comparisons are shown in Table 21. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (3) = 60.032, p < .001. There was thus no evidence that the Bai-Mandarin bilingual adults varied pitch minimum to distinguish narrow focus from broad focus in their Mandarin, regardless of tone.

Table 21. Pitch minimum, narrow focus (NF-m) vs. broad focus (BF), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	p
0 null (speaker, sentence, random intercept only)	4	−707.29				
1 + tone	7	−677.28	0 vs 1	60.032	3	.000***
2 + focus condition	8	−677.20	1 vs 2	0.154	1	.695
3 + tone : focus condition	11	−676.29	2 vs 3	1.826	3	.609

5.3.5.3 Effect of contrastiveness: contrastive focus (CF-m) vs. narrow focus (NF-m)

The results of the fit and model comparisons are shown in Table 22. As can be seen, the best-fit model was Model 1, which contained only the main effect of tone, χ^2 (3) = 62.658, p < .001. There was thus no evidence that the Bai-Mandarin bilingual adults varied pitch minimum to distinguish contrastive focus from narrow focus (i.e., non-contrastive focus) in their Mandarin, regardless of tone.

Table 22. Pitch minimum, contrastive focus (CF-m) vs. narrow focus (NF-m), fit and comparison of nested models, "$\Delta\chi^2$" indicates the change of chi-square, and "Δdf" indicates the change of degrees of freedom.

Models	N_{pars}	−2 LLR	Comparison			
			Models	$\Delta\chi^2$	Δdf	*p*
0 null (speaker, sentence, random intercept only)	4	−699.69				
1 + tone	7	−668.36	0 vs 1	62.658	3	.000***
2 + focus condition	8	−668.31	1 vs 2	0.087	1	.768
3 + tone : focus condition	11	−667.50	2 vs 3	1.622	3	.655

5.3.6 Interim summary

To summarize, we found that the Bai-Mandarin early bilingual adults used duration and pitch-related cues to encode narrow focus in their Mandarin, regardless of tone. Specifically, the bilingual adults lengthened the duration of the focal constituents in the narrow focus condition in comparison to their counterparts in the pre- and post-focus conditions in all the lexical tones. However, there was no evidence that duration was used to distinguish focus types that differed in the size and contrastiveness of the focal constituent, regardless of tone. Regarding the use of pitch-related prosodic cues, we found that the bilinguals expanded the pitch span and raised the pitch maximum for distinguishing narrow focus from post-focus in all the lexical tones. However, there was thus no evidence that pitch-related prosodic cues were used for distinguishing narrow focus from pre-focus, regardless of tone. In addition, the bilinguals did not vary pitch-related prosodic cues for encoding focus types that differed in the size and contrastiveness of the constituent, regardless of tone.

5.4 Discussion and conclusion

The present study examined prosodic focus-marking in Mandarin produced by Bai-Mandarin early bilingual adults who had formally started their Standard Mandarin education at the age of six. Informally, they had exposed to Dali Mandarin (a regional variety of Mandarin) in their daily life before the age of six. Further, they received intensive training in Standard Mandarin, and had been actively using Standard Mandarin in their daily lives. We aimed to establish the ultimate attainment of prosodic focus-marking in Bai-Mandarin early bilinguals' Mandarin. Our results do not provide evidence for the hypothesis that the Bai-Mandarin early bilingual adults under investigation have obtained native-like competence in prosodic focus-marking in Mandarin. We have found that the Bai-Mandarin early bilinguals are highly proficient in using duration and pitch-related prosodic cues to encode focus in Mandarin**,** but they are not fully Standard-native-like.

Regarding the use of prosodic cues for encoding narrow focus, we found that the bilinguals lengthen the duration of the focal constituent in comparison to its counterpart in the pre- and post-focus conditions, like monolingual speakers of Standard Mandarin (Xu, 1999; Yang, 2017; Yang & Chen, 2014). In addition, we found that the bilinguals expanded the pitch span and raised the pitch maximum of the focal constituent to distinguish narrow focus from post-focus in their L2-Mandarin, like monolingual speakers of Standard Mandarin (Yang, 2017; Yang & Chen, 2014). However, the bilinguals do not vary pitch-related prosodic cues, including pitch span, pitch maximum, and pitch minimum to distinguish narrow focus from pre-focus, unlike monolingual speakers of Standard Mandarin.

As for the use of prosodic cues for encoding focus types that differed in the size of the focal constituent, we found that the bilinguals vary neither duration nor pitch-related prosodic cues to distinguish narrow focus from broad focus, unlike monolingual speakers of Standard Mandarin in semi-spontaneous speech (Yang, 2017; Yang & Chen, 2014). Specifically, monolingual speakers of Standard

Mandarin vary duration for distinguishing focus types that differed in the size of the focal constituent.

Regarding the use of prosodic cues for encoding focus types that differed in the contrastiveness of the focal constituent, there is no evidence that the bilinguals vary pitch-related cues or duration to distinguish narrow focus from contrastive focus, similar to monolingual speakers of Standard Mandarin in semi-spontaneous speech (Yang, 2017; Yang & Chen, 2017).

Thus, similar to Tsat-Mandarin early bilinguals (Wang et al., 2012) and QZSM-Mandarin early bilinguals (Chen et al., 2014), the Bai-Mandarin bilinguals' ability to use duration to encode narrow focus is Mandarin-native-like. However, their use of pitch-related cues to differentiate narrow focus from pre-focus, and differentiate focus types is not fully Standard Mandarin native-like. Such non-native manipulation of pitch for encoding focus is also observed in Tsat-Mandarin and older QZSM-Mandarin early bilinguals.

Taken together, we suggest that the different degrees of success in acquiring the use of duration and pitch in early bilinguals' L2 reviewed above might be explained by the similarities and differences between their L1 and L2 in prosodic focus-marking. As reviewed in the introduction, only duration is used for encoding focus in Tsat-Mandarin, QZSM-Mandarin and Bai-Mandarin bilinguals' L1. It seems that the similarity between bilinguals' L1 and L2 can facilitate the learning of the duration cue. Furthermore, although our Bai-Mandarin early bilinguals started their formal Standard Mandarin education at the age of six, they had been exposed to non-standard Mandarin (i.e., Dali Mandarin) informally before the age of six years. It is known that duration is exploited in Dali Mandarin for encoding focus (Chapter 3), which might facilitate the acquisition of using duration for encoding focus in our Bai-Mandarin early bilinguals'L2 Mandarin as well. However, unlike Tsat-Mandarin bilinguals but similar to QZSM-Mandarin bilinguals, our Bai-Mandarin early bilinguals have shown a remarkable ability to expand the pitch span and to raise the pitch maximum for distinguishing focus from post-focus in their Mandarin.

In line with Chen et al. (2014), we suggest that the intensive training in Standard Mandarin might account for the Bai-Mandarin early bilinguals' success in varying pitch-related cues for encoding focus, although they are still not fully native-like.

In future research, more insights into the ultimate attainment in early bilinguals' prosodic focus-marking can be obtained by investigating how monolingual speakers of Standard Mandarin perceive the degree of nativeness of the prosodic encoding of focus in the Mandarin produced by Bai-Mandarin early bilinguals. The second issue for further study is the link between production and comprehension of prosodic focus-marking in Bai-Mandarin early bilinguals' Mandarin. Previous studies of L2 learners have shown that in some cases, the production of certain grammatical features may precede comprehension in L2 learners (Tasseva-Kurktchieva, 2015). Only a few studies have examined perception, differentiated from comprehension, of prosodic focus-marking in bilinguals (Wang et al., 2012; Xu et al., 2012). For example, Xu et al. (2012) used SVO utterances recorded in a production experiment to test whether Mandarin-Taiwanese bilinguals could judge which of the constituents, if any, was emphasized. However, it is questionable whether the "emphasized constituent" perceived by the listeners can be seen as an indication of an established link between perceptual salience and information packaging. If Bai-Mandarin early bilingual's production of prosodic focus-marking goes hand in hand with their comprehension, we are confident in concluding that early bilinguals can ultimately attain native-like competence in their L2. It has been suggested that the production of focus-to-accentuation mapping does not necessarily go hand in hand with comprehension in monolingual adult speakers (Lentz & Chen, 2015). If there is evidence for a production-precedes-comprehension asymmetry or a comprehension-precedes-production asymmetry in early bilinguals' L2, it would imply that ultimate attainment can differ across different modalities in an L2.

CHAPTER 6 General Discussion and Conclusions

6.1 Introduction

How bilinguals master the art of encoding focus prosodically is a long-standing topic of interest for researchers of language development in different disciplines. Previous studies primarily focused on the acquisition of prosodic focus-marking in monolinguals, and generally agreed that monolingual children undergo a complicated developmental process to use prosody appropriately for encoding focus (Arnhold, 2016; Chen, 2009, 2011; Grigos & Patel, 2010; Müller et al., 2006; Romøren, 2016; Wonnacott & Watson, 2008; Yang & Chen, 2014). However, research on prosodic focus-marking in a bilingual context has primarily focused on adults' competence rather than on the developmental aspect (Barnes & Michnowicz, 2015; Bullock, 2009; Colantoni, 2011; Colantoni & Gurlekian, 2004; Grosser, 1997; Gut & Pillai, 2014; O'Rourke, 2005, 2012; Swerts & Zerbian, 2010; Van Rijswijk & Muntendam, 2012; Van Rijswijk et al., 2017; Zerbian, 2013). This thesis investigated how Bai-Mandarin early bilingual children aged six to thirteen acquire prosodic focus-marking in L2-Mandarin.

Central to the present study was the question of how early bilinguals acquired prosodic focus-marking in their L2. More specifically, we aimed to gain insight into two key issues: (1) Does early bilingual children's L2 acquisition show the same rate and route of acquisition as monolingual children's L1 acquisition? (2) How do L1 and L2-input shape the early bilingual children's acquisition of prosodic focus-marking? To this end, we examined the use of prosody for encoding narrow focus and focus types that differed in constituent size and contrastiveness in Bai produced by (near-) monolingual speakers of Bai (Research question 1); Dali Mandarin produced by (near-) monolingual speakers of Dali Mandarin (Research question 2); Mandarin produced by Bai (L1)-Mandarin (L2) early bilingual children aged from six to thirteen years old (Research question 3); and Mandarin produced by Bai (L1)-Mandarin (L2) early bilingual adults (Research question 4).

In the following sections, we first summarise the main findings of each investigation presented in Chapters 2, 3, 4, and 5 and discuss how the findings supported or rejected the hypotheses and predictions proposed in Chapter 1 (Section 6.2), and then discuss the two key issues raised in Chapter 1 (Section 6.3). Last, we suggest directions for future studies of bilinguals' prosodic development (Section 6.4).

6.2 Main findings and hypotheses revisited

Regarding the question of what the role of prosody in focus-marking in Bai is (Research question 1), considering that the use of pitch and duration for encoding focus in tone languages is not related to the number of lexical tones (Jannedy, 2007; Wang et al., 2011; Xu, 1999) and the lack of consistent evidence for the influence of the prestigious language on the less prestigious language in a language contact situation in prosodic focus-marking (Wang et al., 2012, 2011; Xu et al., 2012), we proposed two opposing hypotheses (Hypothesis 1a and 1b). The hypotheses and corresponding predictions are repeated in (1):

(1) Hypothesis 1a: the manner of using duration and pitch for encoding focus in Bai is not similar to that in Standard Mandarin.

Predictions: (near-) monolingual speakers of Bai will only lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and they will vary neither duration nor pitch to distinguish focus types that differed in constituent size and contrastiveness in Bai, as found in other tone languages spoken in the south of China.

Hypothesis 1b: the manner of using duration and pitch for encoding focus in Bai is similar to that in Standard Mandarin.

Predictions: (near-) monolingual speakers of Bai will expand the pitch span and lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and they

> will vary duration and pitch to distinguish focus types that differed in constituent size and contrastiveness in Bai, as found in Standard Mandarin.

We have found that (near-) monolingual speakers of Bai lengthen the duration of the focal constituent to distinguish narrow focus from post-focus. However, they do not vary pitch-related cues to encode narrow focus. Furthermore, they use neither pitch-related cues nor duration to differentiate focus types (narrow focus, contrastive focus, and broad focus). Thus, the prosodic realization of focus in Bai is different from that in Standard Mandarin. As found in Yang and Chen (2014) and Yang (2017), monolingual speakers of Standard Mandarin mark narrow focus prosodically by expanding the pitch span and lengthening the duration of the focal constituent, although they only vary duration to differentiate narrow focus from broad focus. Compared to speakers of Standard Mandarin, speakers of Bai use prosody to a much lesser degree. Our results thus confirmed Hypothesis 1a that Bai is different from Standard Mandarin in the use of prosody for focus-marking purpose, contra Hypothesis 1b. This finding implies that Bai does not appear to be subjected to influence from Standard Mandarin in prosodic focus-marking, despite the fact that Bai has been in long-lasting and intensive contact with Mandarin and has shown lexical and syntactic influence from Mandarin. Bai is thus similar to Yi, Deang, Taiwanese and Tsat in this respect.

Regarding the question of how prosody is used to encode focus in Dali Mandarin (Research question 2), given that varieties of Mandarin spoken in the vicinity of a language other than Mandarin have shown different uses of prosody for marking focus compared to Standard Mandarin, such as Taiwan Mandarin (Xu et al., 2012), we proposed Hypothesis 2. The hypothesis and corresponding predictions are repeated in (2).

> (2) Hypothesis 2: Dali Mandarin, which has been in long-lasting and intensive contact with Bai, exploits prosody for encoding focus in a similar manner as Bai.

> Predictions: (near-) monolingual speakers of Dali Mandarin will lengthen the duration of the focal constituent relative to the non-focal counterpart for encoding narrow focus, and they will vary neither pitch-related cues to encode narrow focus nor duration to encode focus types that differed in constituent size, like speakers of Bai.

We have found that (near-) monolingual speakers of Dali Mandarin lengthen the duration of the focal constituent relative to its counterpart in the post-focus and pre-focus condition, but they do not vary pitch-related cues to encode narrow focus. Dali Mandarin thus differs from Standard Mandarin but resembles Bai in the manner of prosodically realizing narrow focus. Furthermore, (near-) monolingual speakers of Dali Mandarin vary neither pitch-related cues nor duration to differentiate focus types that differed in constituent size, similar to (near-) monolingual speakers of Bai (Chapter 2) but different from monolingual speakers of Standard Mandarin (Yang, 2017). Finally, (near-) monolingual speakers of Dali Mandarin vary neither duration nor pitch-related prosodic cues to differentiate focus types that differed in the contrastiveness of the focal constituent, similar to (near-) monolingual speakers of Bai (Chapter 2) and monolingual speakers of Standard Mandarin (Yang, 2017). Thus, Dali Mandarin resembles Bai in only exploiting duration for encoding narrow focus, instead of Standard Mandarin. Taken together, our results confirmed Hypothesis 2 that Dali Mandarin, which has been in long-lasting and intensive contact with Bai, exploits prosody for encoding focus in a similar manner as Bai. Unlike many regional varieties of Mandarin and Standard Mandarin (Chen, 2010; Duan & Jia, 2014; Duàn et al., 2013; Ouyang & Kaiser, 2015; Shen & Xu, 2016, 2016; Wang et al., 2011; Xu, 1999; Yang, 2017), focus is prosodically encoded via duration in Dali Mandarin. Together with findings on Taiwan Mandarin (Chen et al., 2009; Xu et al., 2012), our finding shows that close contact with a less prestigious language can lead to remarkable differences in prosodic focus-marking between regional varieties of Mandarin and Standard Mandarin.

Regarding the question of what the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children's Mandarin is (Research question 3), as duration might be easier to master relative to pitch due to the limited acoustic space available for focus-related manipulation in pitch for learners of tone languages (Yang & Chen, 2017), we proposed Hypothesis 3a, which is repeated in (3) with its corresponding predictions. On the other hand, as only duration is used to distinguish focus types whereas both duration and pitch-related cues are used to a similar extent in distinguishing narrow focus from non-focus in (semi-) spontaneous speech in Standard Mandarin (Yang, 2017), it is possible that the use of prosody in distinguishing focus types is easier to master than in distinguishing narrow focus and non-focus for learners of Standard Mandarin. This is indeed what has been found for monolingual Mandarin-speaking children (Yang & Chen, 2017). Given that Bai-Mandarin early bilingual children's L2 formal target is Standard Mandarin, and that early bilingual children's L2 acquisition can resemble monolingual children's L1 acquisition in some linguistic domains (Dulay & Burt, 1973, 1974; Jia, 2003; Krashen, 1982; Paradis, 2005, 2007), we proposed Hypothesis 3b. The hypotheses (Hypothesis 3a and 3b) and corresponding predictions are repeated in (3).

> (3) Hypothesis 3a: Bai-Mandarin early bilingual children will learn the use of duration earlier than the use of pitch-related cues for focus-marking purposes.
>
> Prediction: Bai-Mandarin early bilingual children will lengthen the duration of the focal constituent to distinguish narrow focus from pre- and post-focus in all lexical tones, and use pitch-related cues for the same purpose only in some tones.
>
> Hypothesis 3b: Bai-Mandarin early bilingual children will learn the use of prosody in distinguishing focus types earlier than in distinguishing narrow focus and non-focus.

> Prediction: Bai-Mandarin early bilingual children will master the use of duration to distinguish narrow focus from broad focus in all lexical tones earlier than the use of duration and pitch-related cues for distinguishing narrow focus from non-focus in all lexical tones.

We studied the developmental trajectory of prosodic focus-marking in Bai-Mandarin early bilingual children's L2-Mandarin by examining Bai-Mandarin early bilingual children who started their formal Standard Mandarin education at the age of six. Three age groups were examined: six- to seven-year-old, nine- to ten-year-old, and twelve- to thirteen-year-old bilinguals. We have found that Bai-Mandarin early bilingual children acquire the use of duration in encoding narrow focus at the age of six to seven years. Specifically, children of all the three age groups lengthen the duration of the focal constituent to distinguish narrow focus from pre- and post-focus in all lexical tones in their L2-Mandarin, like monolingual Mandarin-speaking adults. However, none of the age groups use pitch-related prosodic cues, including pitch span, pitch maximum, and pitch minimum to distinguish narrow focus from non-focus, regardless of tone. Furthermore, they vary neither duration nor pitch-related prosodic cues to differentiate narrow focus from broad focus, regardless of tone. Finally, none of the age groups make use of duration or pitch-related prosodic cues for distinguishing contrastive focus from non-contrastive focus, regardless of tone. Our results thus confirmed Hypothesis 3a that Bai-Mandarin early bilingual children acquire the use of duration earlier than the use of pitch-related cues for focus-marking purposes in their Mandarin. In line with Yang and Chen (2017) and Yang (2017), we believe that it is easier to master the use of duration relative to pitch as a prosodic cue for focus-marking purposes due to the limited acoustic space available for focus-related manipulation in pitch for learners of tone languages at an early stage of their language acquisition. Further, as Bai-Mandarin early bilingual children's L1-Bai exploits duration for encoding narrow focus, we suggest that the use of duration as a prosodic cue for encoding narrow focus in bilinguals' L1 facilitates the acquisition of using duration for encoding narrow focus in bilinguals' L2. However, our results did not confirm Hypothesis 3b and show that Bai-Mandarin early bilingual children have not acquired the use of prosody in

distinguishing focus types earlier than in distinguishing narrow focus and non-focus. As Bai-Mandarin early bilingual children's L1-Bai uses neither pitch nor duration to encode focus types that differed in constituent size, we suggest that the unsuccessful acquisition of prosodic cues for encoding focus types in the bilingual children's L2 might be explained by negative L1 transfer. Further, the non-Standard Mandarin-like input might also play an important role; we postpone further discussion on this till the following section.

Regarding the question of what the ultimate attainment of prosodic focus-marking in early bilinguals' L2 is (Research question 4), we examined prosodic focus-marking in Mandarin produced by Bai-Mandarin bilingual adults who were teachers of Standard Mandarin at a primary school. These speakers had formally acquired Standard Mandarin at an early age (i.e., six to seven years old). Informally, they had been exposed to Dali Mandarin (a regional variety of Mandarin) in their daily life before the age of six. Further, they had intensive training in Standard Mandarin, used Standard Mandarin actively, and had high Standard Mandarin proficiency. Considering that early exposure to L2, extensive use of L2 and quality of L2 training are crucial to whether early bilinguals can obtain native-like competence in L2 (Chen et al., 2014; Gut & Pillai, 2014; Huang & Jun, 2011; Wang et al., 2012), and that the school teachers could represent the Bai-Mandarin early bilinguals with the highest level of proficiency in Standard Mandarin, we proposed Hypothesis 4. The hypothesis and corresponding predictions are repeated in (4).

> (4) Hypothesis 4: Bai-Mandarin early bilingual adults can achieve native-like competence in prosodic focus-marking in Mandarin.
>
> Predictions: Bai-Mandarin early bilingual adults will master the use of duration and pitch-related cues to distinguish narrow focus from non-focus in all lexical tones, and they will master the use of duration to distinguish narrow focus from broad focus in all lexical tones, like monolingual speakers of Standard Mandarin.

We have found that Bai-Mandarin early bilingual adults lengthen the duration of the focal constituent to differentiate narrow focus from pre- and post-focus, similar to that in monolingual Mandarin-speaking adults' production. Furthermore, they expand the pitch span and raise the pitch maximum of the focal constituent to differentiate narrow focus from post-focus, while monolingual Mandarin-speaking adults exploit pitch-related cues and duration to differentiate narrow focus from both pre- and post-focus (Yang, 2017). However, unlike monolingual Mandarin-speaking adults, our Bai-Mandarin early bilingual adults use neither pitch-related cues nor duration for encoding focus types that differed in constituent size and contrastiveness. Thus, our results did not confirm Hypothesis 4 that Bai-Mandarin early bilingual adults can achieve native-like competence in prosodic focus-marking in Standard Mandarin. We have established that Bai-Mandarin early bilingual adults ultimately attained the ability to vary pitch and duration for encoding focus, but not in a fully Standard Mandarin-like way. Similar to Tsat-Mandarin early bilinguals (Wang et al., 2012) and QZSM-Mandarin early bilinguals (Chen et al., 2014), Bai-Mandarin early bilinguals' ability to use duration to encode narrow focus in Mandarin is Standard Mandarin-like. However, their use of pitch-related cues to differentiate narrow focus from pre-focus, and between focus types is not fully Standard Mandarin-like. Such non-native manipulation of pitch for encoding focus is also observed in Tsat-Mandarin and some QZSM-Mandarin early bilinguals. We suggest that the different degrees of success in acquiring the use of duration and pitch in early bilinguals' L2 reviewed above might be explained by the similarities and differences between their L1 and L2 in prosodic focus-marking. As we reviewed in the introduction, only duration is used for encoding focus in Tsat-Mandarin, QZSM-Mandarin and Bai-Mandarin bilinguals' L1. It seems that this similarity between bilinguals' L1 and L2 can facilitate the learning of the duration cue in Mandarin. However, unlike Tsat-Mandarin bilinguals but similar to QZSM-Mandarin bilinguals, our Bai-Mandarin early bilinguals have shown a remarkable ability to expand the pitch span and raise the pitch maximum of the focal constituent for distinguishing narrow focus from post-focus in their Mandarin. In line with Chen et al. (2014), we suggest that the intensive training in Standard Mandarin might

account for the Bai-Mandarin early bilinguals' ability to vary pitch-related cues for encoding narrow focus. We will discuss why Bai-Mandarin bilinguals are not fully Standard Mandarin-like in the following section.

6.3 General discussion and conclusions

Answers to Research question 1 have provided first-hand information of prosodic focus-marking in Bai-Mandarin early bilinguals' L1-Bai. Further, answers to Research question 2 and 4 have provided first-hand information on the characteristics of L2-input in Bai-Mandarin bilingual children's L2 development in terms of prosodic focus-marking. In addition, answers to Research question 3 and 4 have established the developmental trajectory and ultimate attainment of prosodic focus-marking in Bai-Mandarin early bilingual children and adults.

In light of the answers to the four research questions, we can address the similarities and differences in the rate and route of acquisition between early bilinguals' L2 and monolinguals' L1 acquisition, and discuss the influence of L1 and L2-input on the acquisition of prosodic focus-marking in early bilinguals' L2 development. In Table 1, we summarize the above-reviewed results and results from monolingual Mandarin-speaking children (Yang & Chen, 2017) and adults (Yang, 2017).

Table 1. Overview of prosodic focus-marking in Mandarin (produced by monolingual and bilinguals), Dali Mandarin and Bai[34]

Language	Age of Monolingual	Prosodic Cue	Focus vs. Pre-focus		Focus vs. Post-focus		Focus vs. Broad		Focus vs. Contrastive		Age of Bilingual
			Mo-	Bi-	Mo-	Bi-	Mo-	Bi-	Mo-	Bi-	
Mandarin	4-5	Duration	+	+	+	+	+	-	-	-	6-7
		Pitch span	-	-	-	-	-	-	-	-	
		Pitch-max	-	-	> (T1,T2,T4)	-	+	-	-	-	
		Pitch-min	< (T2)	-	> (T1)	-	-	-	-	-	
	7-8	Duration	+	+	+	+	+	-	-	-	9-10
		Pitch span	+	-	> (T2,T4)	-	-	-	-	-	
		Pitch-max	> (T1,T4)	-	> (T1,T2,T4)	-	-	-	-	-	
		Pitch-min	< (T2,T3)	-	> (T1)	-	-	-	-	-	
	10-11	Duration	+	+	+	+	+	-	+	-	12-13
		Pitch span	+	-	> (T2,T3,T4)	-	+	-	-	-	
		Pitch-max	+	-	> (T1,T2,T4)	-	-	-	-	-	
		Pitch-min	+	-	> (T1)	-	-	-	-	-	
	Adult	Duration	+	+	+	+	+	-	-	-	Adult
		Pitch span	+	-	+	+	-	-	-	-	
		Pitch-max	-	-	+	+	-	-	-	-	
		Pitch-min	+	-	> (T1)	-	-	-	-	-	
Dali Mandarin	Adult	Duration	+		+		-		-		
		Pitch span	-		-		-		-		
		Pitch-max	-		-		-		-		
		Pitch-min	-		-		-		-		
Bai	Adult	Duration	-		+		-		-		
		Pitch span	-		-		-		-		
		Pitch-max	-		-		-		-		
		Pitch-min	-		-		-		-		

[34] In Table 1, "+" indicates that the corresponding prosodic cue, such as duration, pitch span, pitch maximum or pitch minimum is used for distinguishing different focus conditions. "-" indicates that there is no evidence that the corresponding prosodic cue is used for differentiating different focus conditions. "T1", "T2", "T3", or "T4" indicates that the use of prosodic cues for distinguishing focus conditions is found in the indicated lexical tones; and "<" or ">" indicates the direction of the differences between different focus conditions. Same notation is also used in Chapter 4, Table 25.

As shown in Table 1, there are both similarities and differences in the route and rate of acquisition of prosodic focus-marking between Bai-Mandarin early bilingual children's L2 and Mandarin-speaking monolingual children's L1. Regarding the developmental route, compared to monolingual Mandarin-speaking children, the bilingual children are similar to monolingual children in their earlier mastery of duration for encoding narrow focus relative to pitch in their L2-Mandarin acquisition. Specifically, monolingual Mandarin-speaking children master the use of duration for encoding narrow focus in a Mandarin-speaking adult-like way in all the lexical tones at the age of four to five years, while the use of pitch-related cues for the same purpose is only present in some tones. Similar to monolingual Mandarin-speaking children who acquired the use of duration earlier than pitch for encoding narrow focus, Bai-Mandarin bilingual children master the use of duration for distinguishing narrow focus from non-focus (i.e., pre- and post-focus) in their Mandarin at the age of six to seven years, but not the use of pitch-related cues for encoding narrow focus. However, different from monolingual children, the bilingual children have not shown earlier mastery of duration in distinguishing focus types than in distinguishing narrow focus and non-focus. Specifically, at the age of four to five years, monolingual Mandarin-speaking children master the use of duration for distinguishing narrow focus from broad focus in a Mandarin-speaking adult-like way but do not use pitch-related cues in an adult-like way. At the age of seven to eight years, they stop using pitch-related cues for differentiating narrow focus from broad focus, similar to monolingual Mandarin-speaking adults. At the age of ten to eleven years, they vary pitch span for distinguishing narrow focus from broad focus, different from monolingual Mandarin-speaking adults in semi-spontaneous speech, but similar to monolingual adult speakers in read speech (Xu, 1999). However, at the age of ten to eleven years, they still have not acquired the use of pitch-related cues to distinguish narrow focus and non-focus in all the lexical tones. Different from Mandarin-speaking children who acquire the use of prosody in distinguishing focus types earlier than in distinguishing narrow focus and non-focus, Bai-Mandarin bilingual children vary neither duration nor pitch in distinguishing focus types at the age of six to seven years. However, by the same age (i.e., six to seven years old), the

bilingual children have mastered the use of duration for distinguishing narrow focus from non-focus (i.e., pre- and post-focus) in their Mandarin.

Regarding the developmental rate, we found differences in the rate of acquisition between early bilinguals' L2 and monolinguals' L1. Bai-Mandarin early bilingual children have not developed similar competence at the age of twelve to thirteen years (i.e., after five years of formal Standard Mandarin education) to four- to five-year-old monolingual Mandarin-speaking children with a similar length of exposure to Standard Mandarin. That shows that bilinguals are relatively slower in the rate of acquisition. Specifically, our results show that early bilingual children have difficulties in varying pitch-related prosodic cues for encoding narrow focus in their Mandarin at the age of six to seven years. Such difficulties have not been overcome in any of the lexical tones either at the age of nine to ten years (i.e., after two years of their formal Standard Mandarin education), or at the age of twelve to thirteen years (i.e., after five years of their formal Standard Mandarin education), although early bilingual adults can use pitch-related cues to some degree. However, monolingual Mandarin-speaking children show their gradual development from limited to fully target-/adult-like ability to vary pitch for encoding narrow focus and focus types that differed in the size of the focal constituent. For example, in terms of using pitch maximum for distinguishing narrow focus from pre-focus, Mandarin-speaking monolingual children have not mastered such use at the age of four to five years; and at the age of seven to eight years, they have mastered such use in Tone 1 and 4. Later, at the age of ten to eleven years, monolingual Mandarin-speaking children show their ability to vary pitch maximum for distinguishing focus from pre-focus in all the lexical tones, like monolingual Mandarin-speaking adults. Contrary to the claims in previous studies in the domains of vocabulary and grammatical acquisition in early bilinguals' L2 that the route and rate of early bilinguals' L2 acquisition are comparable to that of monolingual children's L1 (Dulay & Burt, 1973, 1974; Jia, 2003; Krashen, 1982; Paradis, 2005, 2007), but similar to the claims in previous studies in the domains of segmental phonology acquisition in early bilinguals' L2 (Baker & Trofimovich, 2005; Trofimovich & Baker, 2007; Tsukada

et al., 2004), we found that bilinguals are not comparable to monolinguals, and are relatively slower in their L2 development.

There are two likely causes for the relatively slow rate of acquisition in Bai-Mandarin early bilingual children's L2. First, the difference between Bai (L1) and Mandarin (L2) in prosodic focus-marking may explain the difference in the rate of acquisition. The use of pitch for lexical purposes is present both in Bai-Mandarin bilinguals' L1 (Bai) and L2 (Mandarin), while the use of pitch for prosodic focus-marking at the sentence level is only present in their L2 (Mandarin). As a result, the acquisition of language-specific use of prosody for encoding focus in their L2 might be a challenging task for Bai-Mandarin early bilinguals.

Second, the non-Standard Mandarin-like input may play a role in the relatively slow rate of acquisition in Bai-Mandarin early bilinguals' L2-Mandarin. Monolingual Mandarin-speaking children who live and grow up in Beijing (i.e., the ones included in Yang and Chen, 2017), are not only exposed to Beijing Mandarin but also to Standard Mandarin in everyday communication and via mass media. However, Bai-Mandarin early bilinguals grow up in an area relatively isolated from Standard Mandarin in the southwest of China. While they have receptive access to Standard Mandarin via mass media, they are mainly exposed to non-standard varieties of Mandarin in everyday communication. These non-standard varieties of Mandarin include two varieties of Mandarin spoken in Dali: one is a learner variety of Mandarin (Semi-Standard Mandarin) produced by Bai-Mandarin early bilingual adults; and the other is a regional variety of Mandarin (i.e., Dali Mandarin) produced by (near-) monolinguals of Dali Mandarin, who live in a Bai-speaking area (i.e., Dali) but do not speak Bai. We have shown that neither of these varieties of Mandarin mark focus in a Standard Mandarin-like manner. The influence of the quality of input has been suggested in past work on the development in different linguistic domains of monolinguals (Grünloh et al., 2015; Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989; Snow, 1972), early bilingual adults (Chen, 2014; Chen et al., 2014; Gut & Pillai, 2014), and late bilinguals (Flege, 2009; Flege & Liu, 2001;

Piske et al., 2001). Previous studies of early bilingual adults' competence in prosodic focus-marking only speculated that the unsuccessful acquisition of prosodic focus-marking in early bilinguals' L2 might be explained by the "non-standard-like input" (Chen et al., 2014; Gut & Pillai, 2014). However, the present study closely examined the various strands of L2-input in early bilingual children's language environment. In doing so, it has provided for the first time direct empirical evidence that the quality of L2-input might also influence the rate of acquisition in early bilingual children's L2 development. The non-standard input may in turn contribute the non-native like use of pitch in Mandarin by Bai-Mandarin early bilingual adults with high proficiency in Mandarin.

To conclude, by investigating Bai-Mandarin early bilingual children aged from six to thirteen and bilingual adults, we have established the developmental trajectory and ultimate attainment of prosodic focus-marking in Bai-Mandarin early bilinguals' Mandarin. We have found both similarities and differences in the route and rate of acquisition of prosodic focus-marking between early bilingual children's L2 and monolingual children's L1. Furthermore, we have found that Bai-Mandarin early bilingual adults are highly proficient in using duration and pitch-related prosodic cues for encoding focus in Mandarin**,** but they are not fully Standard Mandarin-like. Our results show that L1 influence (i.e., positive and negative transfer) is evident in the bilinguals' L2 development, which has been widely observed in bilingual language acquisition in different linguistic domains (Baker & Trofimovich, 2005; Barnes & Michnowicz, 2015; Bullock, 2009; Flege & Fletcher, 1992; Simonet, 2008; Tsukada et al., 2004; Wang et al., 2012). Importantly, non-Standard L2 input might also influence the route and rate of acquisition in early bilinguals' L2 development.

One possible criticism on this study is that the numbers of the child participants in the experiments reported in chapters 2 to 5 is small from the perspectives of statistical power and the generalisability of the results, although they are comparable to the published studies of early bilinguals' language development and are larger than the number of speakers used in most studies of prosodic focus-marking in early

bilingual adults. Our data sets had reasonable case-to-condition ratios, which are calculated by dividing the numbers of the cases by the total number of combinations of conditions for each age group and outcome variable separately. Specifically, the case-to-condition ratio ranges from 36:1 to 50:1 in Chapter 2, from 21:1 to 23:1 in Chapter 3, from 15:1 to 31:1 in Chapter 4, and from 20:1 to 22:1 in Chapter 5. These case-to-condition ratios ensure the reliability of our results. However, the use of a small number of speakers can increase the risk of Type I errors (i.e., false positives). It can thus be very meaningful to conduct replication studies with larger sample sizes to check the generalisability of the current results.

6.4 Suggestions for further research

Our study suggests three topics for future research. One is the further exploration of the developmental trajectory of prosodic focus-marking in bilinguals aged from thirteen to adulthood. In the present study, we have found that the use of pitch-related prosodic cues for encoding focus is absent in Bai-Mandarin early bilingual children's Mandarin at the age of twelve to thirteen years. However, pitch span and pitch maximum are used for encoding narrow focus in Bai-Mandarin early bilingual adults. It indicates that pitch-related prosodic cues can be acquired at a later stage of early bilinguals' L2 development. As Grigos and Patel (2010) suggested, children are likely to continue refining articulator movement to mark focus throughout adolescence. In other words, in future studies, we can explore whether and how the language development of early bilingual children's L2 continues throughout adolescence, and early adulthood. This line of research can bridge the gap in our knowledge of the development after puberty.

The second issue for future study is to determine how monolingual speakers of Standard Mandarin perceive the degree of nativeness of prosodic encoding of focus in Mandarin produced by Bai-Mandarin early bilinguals. If the non-Standard-native-like prosodic focus-marking we observed in Bai-Mandarin bilinguals' Mandarin can influence the perceived nativeness of their L2-Mandarin, this would suggest that

prosodic focus-marking in speech production can be used as a criterion for assessing L2-Mandarin language proficiency.

The third issue for further study is the link between production and comprehension of prosodic focus-marking in Bai-Mandarin early bilinguals' Mandarin. Previous perception studies of prosodic focus-marking simply used an identification task to test utterances produced in a production experiment. For example, Xu (2012) used SVO utterances recorded in a production experiment to test whether speakers of Standard Mandarin judge which of the constituent, or none of the constituents in the utterances, was emphasized. This kind of identification task was widely used in previous investigations of prosodic focus-marking, such as in Northern Sotho (Zerbian, 2007), Bengali (Choudhury & Kaiser, 2012), Standard Mandarin (Xu et al., 2012), Hausa (Hartmann & Zimmermann, 2007), Taiwan Mandarin (Xu et al., 2012), Taiwanese (Xu et al., 2012), and Vietnamese (Jannedy, 2007). These studies showed that listeners could perceive the perceptual salience of the focal constituent. In addition, the post-focus compression (compressed pitch span and intensity of the post-focal constituent) was found to be beneficial to focus recognition rate. For example, Xu (2012) found that the focus recognition rate was substantially lower in Taiwanese and Taiwan Mandarin in comparison to Standard Mandarin, due to the absence of the post-focus compression in Taiwanese and Taiwan Mandarin.[35] However, it is questionable whether the "emphasized constituent" perceived by the listeners can be seen as an indication of an established link between perceptual salience and information packaging. Therefore, comprehension tasks which take the encoded information structure in a given context into consideration, such as Reaction-Time paradigm developed by Birch and Clifton (1995) and adapted for research on language development by Chen (2010, 2014) and the visual-world paradigm based eye tracking task used by Mulders and Szendroj (2016), are needed

[35] Xu et al.(2012) showed that the focus recognition rate in Taiwanese and Taiwan Mandarin was "substantially lower (50.7 – 73.3% correct identification, which is nevertheless well above the chance level of 25% with 4 alternative focus judgements, thanks probably to the phonetic enhancement of the on-focus words, c.f. Wu and Xu 2001 for Cantonese) than in Mandarin (66.7% - 90.9%), and by the fact that even for Mandarin, when such compression is not possible in final focus, the recognition rate (66.7%) is similar to Taiwanese and Taiwan Mandarin."

to shed light on the issue of the comprehension of prosodic focus-marking. It has been suggested that children's production of focus-to-accentuation mapping is similar to their comprehension in L1 development (Chen, 2010). However, for L2 learners, an asymmetry relationship between production and comprehension has been reported in the acquisition of some morpho-syntactic features (Tasseva-Kurktchieva, 2015). As limited studies have examined the link between production and comprehension of prosodic focus-marking in bilingual children, it would be very interesting to know whether the link between bilingual speakers' production and comprehension of prosodic focus-marking is symmetry or asymmetry. In addition, L2 proficiency might play an important role in the link between production and comprehension.

BIBLIOGRAPHY

Allen, B. (2004). *Bai dialect survey*. Kunming: Yunnan Nationalities Publishing House.

Allen, B., Sū, W. (苏玮雅), & Yǐn, M. (尹曼芬). (1997). Báiyǔ xǐzhōuzhèn shēngdiào de cèshì fēnxī [The experimental study on Xizhou Bai lexical tones, 白语喜洲镇话声调的测试分析]. *Dàlǐ Xuéyuàn xuébào[Journal of Dali University: General Edition, 大理学院学报：综合版]*, *2*, 54–58.

Arnhold, A. (2016). Complex prosodic focus marking in Finnish: Expanding the data landscape. *Journal of Phonetics*, *56*, 85–109.

Arnhold, A., Chen, A., & Järvikivi, J. (2016). Acquiring complex focus-marking: Finnish 4- to 5-year-olds use prosody and word order in interaction. *Frontiers in Psychology*, *7*, 1–19.

Asu, E. L., & Nolan, F. (2007). The analysis of low accentuation in Estonian. *Language and Speech*, *50*(4), 567–588.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Backman, N. (1979). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage Studies Bulletin*, 239–265.

Baker, W., & Trofimovich, P. (2005). Interaction of native-and second-language vowel system (s) in early and late bilinguals. *Language and Speech*, *48*(1), 1–27.

Bard, E. G., & Aylett, M. P. (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the International Congress of Phonetic Sciences (ICPhS-14)* (pp. 1753–1756). San Francisco.

Barnes, H., & Michnowicz, J. (2015). Broad focus declaratives in Veneto-Spanish bilinguals: Peak alignment and language contact. *Studies in Hispanic and Lusophone Linguistics*, *8*(1), 35–57.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv:1506.04967 [Stat]*. Retrieved from http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from http://CRAN.R-project.org/package=lme4

Bauer, R. S., Cheung, K., Cheung, P., & Ng, L. (2001). Acoustic correlates of focus-stress in Hong Kong Cantonese. In *Eleventh Annual Meeting of the Southeast Asian Linguistics Society* (pp. 29–49).

Baumann, S., Becker, J., Grice, M., & Mücke, D. (2007). Tonal and articulatory marking of focus in German. In J. Trouvain & J. B. William (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1029–1032). Saarbrücken: Saarland University.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, *11*(1–2), 17–68.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(01), 255–309.

Bialystok, E., & Miller, B. (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition*, *2*(02), 127–145.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*(4), 365–391.

Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 5.4.01). Retrieved from http://www.praat.org/

Bruce, G. (1982). Textual aspects of prosody in Swedish. *Phonetica*, *39*(4–5), 274–287.

Bullock, B. E. (2009). Prosody in contact in French: A case study from a heritage variety in the USA. *International Journal of Bilingualism*, *13*(2), 165–194.

Burdin, R. S., Phillips-Bourass, S., Turnbull, R., Yasavul, M., Clopper, C. G., & Tonhauser, J. (2015). Variation in the prosody of focus in head- and head/edge-prominence languages. *Lingua*, *165*, *Part B*, 254–276.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–55). New York: Academic Press.

Chao, Y. (1930). A system of tone letters. *Le Maitre Phonetique*, (45), 24–27.

Chao, Y. (1968). *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, A. (2017). Get the focus right across languages: Acquisition of prosodic focus-marking in production. In P. Prieto & Esteve-Gibert (Eds.), *Prosodic development*. Amsterdam: Benjamins.

Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. Cláudia Vigário, S. Frota, & M. J. Freitas (Eds.), *Phonetics and phonology: Interactions and interrelations* (Vol. 306, pp. 91–106). Amsterdam: John Benjamins Publishing.

Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua*, *120*(8), 1926–1939.

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, *38*(05), 1055–1083.

Chen, A. (2012). The prosodic investigation of information structure. In M. Krifka & R. Musan (Eds.), *The expression of information structure* (pp. 251–286). Berlin: Mouton de Gruyter.

Chen, A. (2014). Production-comprehension (a) symmetry: Individual differences in the acquisition of prosodic focus-marking. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 423–427). Dublin, Ireland.

Chen, S., Wang, B., & Xu, Y. (2009). Closely related languages, different ways of realizing focus. In *Interspeech* (pp. 1007–1010). Brighton, UK.

Chen, Ying. (2014). *Prosodic realization of focus in second language speech: Effects of language experience* (Doctoral dissertation). University of Oregon.

Chen, Ying, Guion-Anderson, S., & Xu, Y. (2012). Post-focus compression in second language Mandarin. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of Speech Prosody 2012* (pp. 410–413). Shanghai, China.

Chen, Ying, Xu, Y., & Guion-Anderson, S. (2014). Prosodic realization of focus in bilingual production of Southern Min and Mandarin. *Phonetica*, *71*(4), 249–270.

Chen, Yiya. (2010). Post-focus F0 compression—Now you see it, now you don't. *Journal of Phonetics*, *38*(4), 517–525.

Chen, Yiya, & Braun, B. (2006). Prosodic realization of information structure categories in Standard Chinese. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of Speech Prosody 2006*. Dresden, Germany.

Chen, Yiya, & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, *36*(4), 724–746.

Chen, Yiya, Lee, P. P., & Pan, H. (2016). Topic and focus marking in Chinese. In C. Féry & S. Ishihara (Eds.), *The Oxford handbook of information structure* (pp. 733–752). Oxford: Oxford University Press.

Choudhury, A., & Kaiser, E. (2012). Prosodic focus in Bangla: A psycholinguistic investigation of production and perception. In *LSA Annual Meeting Extended Abstracts* (Vol. 3, pp. 25–1).

Clark, H. H., Haviland, S. E., & Freedle, R. (1977). Comprehension and the Given-New contrast. In *Discourse production and comprehension* (Vol. 1, pp. 1–40). Norwood, N.J.: Ablex.

Colantoni, L. (2011). Broad-focus declaratives in Argentine Spanish contact and non-contact varieties. In C. Gabriel & C. Lleó (Eds.), *Intonational phrasing in Romance and Germanic: Cross-linguistic and bilingual studies* (pp. 183–212). Amsterdam: John Benjamins Publishing.

Colantoni, L., & Gurlekian, J. (2004). Convergence and intonation: historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition*, *7*(02), 107–119.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142–2156.

Dài, Q. (戴庆夏), & Lǐ, S. (李绍尼). (1992). Hànyǔ duì báiyǔ de yǐngxiǎng [The influences of Mandarin on Bai, 汉语对白语的影响]. In Q. (戴庆夏) Dài (Ed.), *Hànyǔ yǔ shǎoshù mínzú yǔyán guānxì gàilùn [An introduction to the relationship between Mandarin Chinese and minority languages, 《汉语与少数民族语言关系概论》]*. Beijing: Minzu University of China Press.

Dàlǐ báizú zìzhì zhōu dìfāngzhì biānzhuàn wěiyuánhuì bàngōngshì (大理白族自治州地方志编纂委员会办公室) (Ed.). (2011). *Dàlǐ zhōu niānjiàn (2011) [Dali Annuals (2011), 《大理州年鉴(2011)》]*. Kunming: Yúnnán mínzú chūbǎn shè [Yúnnán mínzú Press, 云南民族出版社].

De Jong, J. (1980). On the treatment of focus phenomena in functional grammar. *GLOT, Leids Taalkundig Bulletin*, *3*, 89–115.

De Ruiter, L. E. (2010). *Studies on intonation and information structure in child and adult German* (Doctoral dissertation). Radboud University, Nijmegen.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(04), 499–533.

Dèng, Y. (邓瑶), & Hé, W. (何稳菊). (2012). Yúnnán Dàlǐ Xǐzhōu B áizú jūmíng yǔyán shēnghuó diàochá [A language use and attitude survey of Bai in Xizhou County Dali Yunnan 云南大理喜洲白族居民语言生活调查]. *Mínzú Fānyì [Journal of Minority Translators 民族翻译]*, (3), 78–83.

Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Duan, W., & Jia, Y. (2014). The typology of focus realization of Northern Mandarin. In *Proceedings of the 9th International Symposium on Chinese Spoken Language Processing* (pp. 492–496). Singapore.

Duàn, W. (段文君), Jiǎ, Y. (贾媛), & Rǎn, Q. (冉启斌). (2013). Shāndōngfāngyán jiāodiǎn yǔyīn shíxiàn de gòngxìng hé chāyìxìng tèzhēng -- yǐ jǐnán, liáochéng, zībó fāngyán wéilì [Internal similarities and differences of phonetic realization of the focus in the Shandong dialect – A case study of the Jinan, Liaocheng and Zibo dialects, 山东方言焦点语音实现的共性和差异性特征——以济南，聊城，淄博方言为例]. *Qīnghuá Dàxué Xuébào (Zìrán Kēxué Bǎn) [Journal of Tsinghua University (Science and Technology), 清华大学学报 (自然科学版)]*, *53*(6), 835–838.

Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, *23*(2), 245–258.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*(1), 37–53.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Los Angeles: SAGE Publication Ltd.

Flege, J. E. (2009). Give input a chance. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Bristol: Multilingual Matters.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America*, *91*(1), 370–389.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*(4), 527–552.

Frey, W. (2006). Contrast and movement to the German prefield. In V. Molnár & S. Winkler (Eds.), *The architecture of focus* (pp. 235–264). Berlin: de Gruyter.

Frota, S. (2002). The prosody of focus: A case-study with cross-linguistic implications. In *Proceedings of Speech Prosody 2002* (pp. 315–318). Aix-en-Provence, France.

Genesee, F. (1989). Early bilingual development: one language or two? *Journal of Child Language*, *16*(1), 161–179.

Goodluck, H. (1986). Language acquisition and linguistic theory. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 49–68). Cambridge: Cambridge University Press.

Grigos, M. I., & Patel, R. (2010). Acquisition of articulatory control for sentential focus in children. *Journal of Phonetics*, *38*(4), 706–715.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 225–254). Mahwah, NJ: Lawrence Erlbaum Associates.

Grosjean, F. (2010). *Bilingual*. Cambridge, MA: Harvard University Press.

Grosser, W. (1997). On the acquisition of tonal and accentual features of English by Austrian learners. In A. James & J. Leather (Eds.), *Second language speech: Structure and process* (pp. 211–228). Berlin: Mouton de Gruyter.

Grünloh, T., Lieven, E., & Tomasello, M. (2015). Young children's intonational marking of new, given and contrastive referents. *Language Learning and Development*, *11*(2), 95–127.

Gu, W., & Lee, T. (2007). Effects of focus on prosody of Cantonese speech - A comparison of surface feature analysis and model-based analysis. In *Proceedings of the International Workshop Paralinguistic Speech' 07* (pp. 59–64). Saarbrücken, Germany.

Guasti, M. T. (2004). *Language acquisition: The growth of grammar*. Cambridge, MA: The MIT Press.

Guion, S. G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition*, *27*(04), 503–533.

Gundel, J. K. (1988). Universals of topic-comment structure. *Studies in Syntactic Typology*, *17*, 209–239.

Gundel, J. K. (1999). On different kinds of focus. In P. Bosch & R. Van de Sandt (Eds.), *Focus: Linguistic, cognitive, and computational perspectives* (pp. 293–305). Cambridge: Cambridge University Press.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, *19*(02), 377–417.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C. (2006). Yucatec Maya tone in sentence perspective. *Poster Presented at LabPhon10, Paris*.

Gussenhoven, C. (2007). Notions and subnotions in information structure. *Acta Linguistica Hungarica*, *55*(3–4), 381–395.

Gussenhoven, C. (2008). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus* (pp. 83–100). Dordrecht: Springer.

Gussenhoven, C., & Teeuw, R. (2008). A moraic and a syllabic H-tone in Yucatec Maya. *Fonología Instrumental: Patrones Fónicos Y Variación*, 49–71.

Gut, U., & Pillai, S. (2014). Prosodic marking of information structure by Malaysian speakers of English. *Studies in Second Language Acquisition*, *36*(02), 283–302.

Halliday, M. A. (1967a). Notes on transitivity and theme in English: Part 1. *Journal of Linguistics*, *3*(01), 37–81.

Halliday, M. A. (1967b). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, *3*(02), 199–244.

Halliday, M., Matthiessen, C. M., & Matthiessen, C. (2004). *An introduction to functional grammar*. London: Routledge.

Hanssen, J., Peters, J., & Gussenhoven, C. (2008). Prosodic effects of focus in Dutch declaratives. In *Proceedings of Speech Prosody 2008* (pp. 609–612). Campinas, Brazil.

Hartmann, Katharina, & Zimmermann, M. (2007). In place-out of place? Focus in Hausa. In K. Schwabe & S. Winkler (Eds.), *On information structure, meaning and form: Generalizing across languages* (pp. 365–403). Amsterdam: John Benjamins Publishing Company.

Hé, W. (何稳菊) (2015). *Xīzhōu Báiyǔ dāngzìdiào shíyàn yánjiū [The mono-syllabic tone of Xizhou dialect experimental study 喜洲白语单字调实验研究]* (Master's thesis). Yunnan University.

He, X., Hanssen, J., van Heuven, V. J., & Gussenhoven, C. (2011). Phonetic implementation must be learnt: Native versus Chinese realization of focus accent in Dutch. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 843–846). Hong Kong, China.

Hefright, B. E. (2011). *Language contact as bilingual contrast among Bái language users in Jiànchuān County, China* (Doctoral dissertation). The University of Michigan.

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics*, *31*(1), 39–62.

Hockett, C. F. (1958). A course in modern linguistics. *Language Learning*, *8*(3–4), 73–75.

Hoffmann, C. (2014). *Introduction to bilingualism*. London: Routledge.

Hoot, B. (2012). *Presentational focus in heritage and monolingual Spanish* (Doctoral dissertation). Univeristy of Illinois at Chicago.

Houwer, A. D., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, *35*(6), 1189–1211.

Huang, B. H., & Jun, S.-A. (2011). The effect of age on the acquisition of second language prosody. *Language and Speech*, *54*(3), 387–411.

İşsever, S. (2003). Information structure in Turkish: the word order–prosody interface. *Lingua*, *113*(11), 1025–1053.

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: The MIT Press.

Jannedy, S. (2007). Prosodic focus in Vietnamese. In S. Ishihara, S. Jannedy, & A. Schwarz (Eds.), *Interdisciplinary studies on information structure* (Vol. 8, pp. 209–230). Postdam: Universitätsverlag Potsdam.

Jannedy, S. (2008). The effect of focus on lexical tones in Vietnamese. In A. Botinis (Ed.), *Proceedings of ISCA Tutorial and Research Workshop On Experimental Linguistics* (pp. 113–116). Athens: ISCA and the University of Athens.

Jia, G. (2003). The acquisition of the English plural morpheme by native Mandarin Chinese-speaking children. *Journal of Speech, Language, and Hearing Research*, *46*(6), 1297–1311.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Jun, S.-A., & Lee, H.-J. (1998). Phonetic and phonological markers of contrastive focus in Korean. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP 98)*. Sydney, Australia.

Kelm, O. R. (1987). An acoustic study on the differences of contrastive emphasis between native and non-native Spanish speakers. *Hispania*, *70*(3), 627–633.

Kohnert, K. (2004). Processing skills in early sequential bilinguals. *Bilingual Language Development and Disorders in Spanish-English Speakers*, 53–76.

Krashen, S. D. (1973). Lateralization, language learning, and the critical period: Some new evidence. *Language Learning*, *23*(1), 63–74.

Krashen, S. D. (1982). Accounting for child-adult differences in second language rate and attainment. In S. D. Krashen, M. Long, & R. Scarcella (Eds.), *Child-adult differences in second language acquisition* (Vol. 2, pp. 202–226). Rowley, MA: Newbury House.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*(3–4), 243–276.

Kügler, F., & Genzel, S. (2012). On the prosodic expression of pragmatic prominence: The case of pitch register lowering in Akan. *Language and Speech*, *55*(3), 331–359.

Kügler, F., & Skopeteas, S. (2007). On the universality of prosodic reflexes of contrast: The case of Yucatec Maya. In *Proceedings of the 16th*

International Congress of Phonetic Sciences (pp. 1025–1028). Saarbrücken.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests in linear mixed effect models. *R Package Version 2.0-20*. Retrieved from http://CRAN.R-project.org/package=lmerTest

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Lakshmanan, U. (1995). Child second language acquisition of syntax. *Studies in Second Language Acquisition*, *17*(03), 301–329.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lanza, E. (1992). Can bilingual two-year-olds code-switch? *Journal of Child Language*, *19*(3), 633–658.

Lenneberg, E. H. (1967). *The biological foundations of language*. New York: John Wiley & Sons.

Lentz, T. O., & Chen, A. (2015). Unbalanced adult production and perception in prosody. In *Proceedings of International Congress of Phonetic Sciences*. Glasgow, UK : University of Glasgow.

Lǐ, L. (李琳). (2009). Xiǎoyì yúnnán xiàguāng fāngyán hé Pǔtōnghuà zhījiān de yǔyīn chābié [On the differences between Yunnan Xiaguang dialect and Putonghua's sound pattern, 小议云南下关方言和普通话之间的语音差别]. *Yúnnán Nóngyè Dàxué xuébào[Journal of Yunnan Agricultural University, 云南农业大学学报]*, *3*(1), 85–88.

Li, Y. (2015). *Language planning in China*. Berlin: Walter de Gruyter.

Lǐ, Y. (李义祝). (2012). *Yúnnán Hèqìn hànyǔ fāngyán hé báiyǔ de yǔyánjiēchù yánjiū [A study on the language contact between Bai and Mandarin in Heqin, Yunnan, 《云南鹤庆汉语方言和白语的语言接触研究》]* (Master's thesis). Jinan University.

Lin, Y.-H. (2007). *The sounds of Chinese*. Cambridge: Cambridge University Press.

Liu, Z., Chen, A., & Van de Velde, H. (2014). Prosodic focus marking in Bai. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 628–631). Dublin, Ireland.

Liu, Z., Chen, A., & Van de Velde, H. (2016a). Prosodic focus marking in Bai-Mandarin sequential bilinguals' Mandarin. In *Proceedings of Speech Prosody 2016* (pp. 951–955). Boston, USA.

Liu, Z., Chen, A., & Van de Velde, H. (2016b). Prosodic focus marking in minority L1 Bai-children learning Mandarin Chinese as L2. In J. Scott & D. Waughtal (Eds.), *Proceedings of the 40th annual Boston University Conference on Language Development*. Boston, USA.

Liu, Z., Van de Velde, H., & Chen, A. (2016). Prosodic focus marking in Dali Mandarin. In C. DiCanio, J. Malins, J. Good, K. Michelson, J. Jaeger, & H. Keily (Eds.), *Proceedings of the 5th International Symposium on Tonal Aspects of Language* (pp. 103–106). Buffalo, New York.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*(03), 251–285.

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, *6*, 1–7.

Man, V. C. H. (2002). Focus effects on Cantonese tones: An acoustic study. In *Proceedings of Speech Prosody 2002* (pp. 467–470). Aix-en-Provence, France.

Maskikit-Essed, R., & Gussenhoven, C. (2016). No stress, no pitch accent, no prosodic focus: the case of Ambonese Malay. *Phonology*, *33*(02), 353–389.

McGory, J. T. (1997). *Acquisition of Intonational Prominence in English by Seoul Korean and Mandarin Chinese Speakers* (Doctoral dissertation). The Ohio State University.

McLaughlin, B. (2012). *Second language acquisition in childhood: Volume 2. School-age children*. New York: Psychology Press.

Meisel, J. M. (2004). The bilingual child. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 90–113). Malden, MA: Blackwell Publishing.

Meisel, J. M. (2008). Child second language acquisition or successive first language acquisition. In B. Haznedar & E. Gavruseva (Eds.), *Current trends in child second language acquisition* (pp. 55–80). Amsterdam: John Benjamins Publishing Company.

Michaud, A., & Brunelle, M. (2016). Information structure in Asia: Yongning Na (Sino-Tibetan) and Vietnamese (Austroasiatic). In C. Féry & S. Ishihara (Eds.), *The Oxford handbook of information structure* (pp. 774–789). Oxford: Oxford University Press.

Michaud, A., & Vu-Ngoc, T. (2004). Glottalized and nonglottalized tones under emphasis: Open quotient curves remain stable, F0 curve is modified. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2004* (pp. 745–748). Nara, Japan.

Mulders, I., & Szendroj, K. (2016). Early association of prosodic focus with alleen "only": Evidence from eye movements in the visual-world paradigm. *Frontiers in Psychology*, *7*, 1–19.

Müller, A., Höhle, B., Schmitz, M., & Weissenborn, J. (2006). Focus-to-stress alignment in 4-to 5-year-old German-learning children. In A. Belletti, E. Bennati, C. Chesi, E. DiDomenico, & I. Ferrari (Eds.), *Proceedings of GALA 2005* (pp. 393–407). Newcastle: Cambridge Scholars Press.

Muntendam, A., & Torreira, F. (2016). Focus and prosody in Spanish and Quechua. In M. E. Armstrong, N. Henriksen, & M. del M. Vanrell (Eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields* (pp. 69–89). Amsterdam: John Benjamins Publishing Company.

Nava, E., & Zubizarreta, M. L. (2008). Prosodic Transfer in L2 Speech: Evidence from Phrasal Prominence and Rhythm. In *Proceedings of Speech Prosody 2008* (pp. 335–338). Campinas, Brazil.

Nava, E., & Zubizarreta, M. L. (2009). Order of L2 acquisition of prosodic prominence patterns: Evidence from L1 Spanish/L2 English speech. In J.

Crawford, K. Otaki, & M. Takahashi (Eds.), *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)* (pp. 175–187). Somerville, MA: Cascadilla Proceedings Project.

Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, *16*(1), 55–68.

O'Brien, M., & Gut, U. (2010). Phonological and phonetic realisation of different types of focus in L2 speech. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 331–336). Poznań, Poland.

Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, *28*(2), 191–230.

O'Rourke, E. (2005). *Intonation and language contact: A case study of two varieties of Peruvian Spanish* (Doctoral dissertation). University of Illinois at Urbana-Champaign.

O'Rourke, E. (2012). The realization of contrastive focus in Peruvian Spanish intonation. *Lingua*, *122*(5), 494–510.

Ouyang, I. C., & Kaiser, E. (2015). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*, *30*(1–2), 57–72.

Paradis, J. (2005). Grammatical morphology in children learning English as a second language: Implications of similarities with specific language impairment. *Language, Speech, and Hearing Services in Schools*, *36*(3), 172–187.

Paradis, J. (2007). Second language acquisition in childhood. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 387–405). Malden: John Wiley & Sons.

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, *18*(1), 41–58.

Penfield, W., & Roberts, L. (2014). *Speech and brain mechanisms*. Princeton, New Jersey: Princeton University Press.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425.

R Core Team. (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Retrieved from http://www.R-project.org/.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, *28*(2007), 41–66.

Rasier, L., & Hiligsmann, P. (2009). Exploring the L1-L2 relationship in the L2 acquisition of prosody. In *Proceedings of First and Second Languages: Exploring the Relationship in Pedagogy-related Context*. Oxford.

Rasier, L., Hiligsmann, P., Caspers, J., & Van Heuven, V. (2010). Accentual marking of information status in Dutch and French as foreign languages. Production and perception. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 379–385). Poznań, Poland.

Romøren, A. S. H. (2016). *Hunting highs and lows: The acquisition of prosodic focus marking in Swedish and Dutch* (Doctoral dissertation). Utrecht University.

Romøren, A. S. H., & Chen, A. (2015). Quiet is the new loud: Pausing and focus in child and adult Dutch. *Language and Speech*, *58*(1), 8–23.

Rump, H. H., & Collier, R. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech*, *39*(1), 1–17.

Schauber, E. (1978). Focus and presupposition: A comparison of English intonation and Navajo particle placement. In D. J. Napoli (Ed.), *Elements of tone, stress, and intonation* (pp. 144–173). Washington, D.C.: Georgetown University Press.

Schwartz, B. D. (2004). Why child L2 acquisition? In J. Van Kampen & S. Baauw (Eds.), *Proceedings of Generative Approaches to Language Acquisition 2003* (pp. 47–66). Utrecht, The Netherlands: LOT Occasional Series.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*(2), 111–123.

Selkirk, E. (1986). *Phonology and syntax: The relationship between sound and structure*. Cambridge, MA: The MIT press.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Cambridge, Massachusetts: Blackwell.

Sgall, P., Hajicová, E., & Panevová, J. (1986). *The meaning of the sentence in its semantic and pragmatic aspects*. Dordrecht: Reidel.

Shen, C., & Xu, Y. (2016). Prosodic Focus with Post-focus Compression in Lan-Yin Mandarin. In *Proceedings of Speech Prosody 2016*. Boston, MA, USA.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, *3*, 83–109.

Simonet, M. (2008). *Language contact in Majorca: An experimental sociophonetic approach* (Doctoral dissertation). University of Illinois at Urbana-Champaign.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, *43*(2), 549–565.

Swerts, M., & Zerbian, S. (2010). Intonational differences between L1 and L2 English in South Africa. *Phonetica*, *67*(3), 127–146.

Tasseva-Kurktchieva, M. (2015). Can production precede comprehension in L2 acquisition? *Second Language Research*, *31*(4), 493–522.

Thomason, S. G. (2001). *Language contact*. Edinburgh: Edinburgh University Press.

Treffers-Daller, J. (1997). Variability in code-switching styles: Turkish-German code-switching patterns. In R. Jacobson (Ed.), *Code-switching Worldwide* (pp. 177–197). Berlin: Mouton de Gruyter.

Trofimovich, P., & Baker, W. (2007). Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics*, *28*(02), 251–276.

Tsukada, K., Birdsong, D., Mack, M., Sung, H., Bialystok, E., & Flege, J. E. (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica*, *61*(2–3), 67–83.

Tucker, G. R. (1998). A global perspective on multilingualism and multilingual education. In J. Cenoz & F. Genesee (Eds.), *Beyong bilingualism: Multilingualism and multilingual education* (pp. 3–15). Clevedon: Multilingual Matters.

Turco, G., Dimroth, C., & Braun, B. (2015). Prosodic and lexical marking of contrast in L2 Italian. *Second Language Research*, *31*(4), 465–491.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff (Ed.), *Methods in empirical prosody research* (pp. 1–28). Berlin: Mouton de Gruyter.

Ueyama, M., & Jun, S.-A. (1996). Focus realization of Japanese English and Korean English intonation. *UCLA Working Papers in Phonetics*, 110–125.

Unsworth, S. (2005). *Child L2, adult L2, child L1: Differences and similarities. A study on the acquisition of direct object scrambling in Dutch* (Doctoral dissertation). Utrecht University.

Vallduví, E. (1990). *The informational component* (Doctoral dissertation). University of Pennsylvania, Philadelphia, USA.

Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, *34*(3), 459–520.

Van Rijswijk, R., & Muntendam, A. (2012). The prosody of focus in the Spanish of Quechua-Spanish bilinguals: A case study on noun phrases. *The International Journal of Bilingualism*, *18*(6), 614.

Van Rijswijk, R., Muntendam, A., & Dijkstra, T. (2017). Focus marking in Dutch by heritage speakers of Turkish and Dutch L1 speakers. *Journal of Phonetics*, *61*, 48–70.

Wang, B., Li, C., Wu, Q., Zhang, X., Wang, B., & Xu, Y. (2012). Production and perception of focus in PFC and non-PFC languages: Comparing Beijing Mandarin and Hainan Tsat. In *INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association* (pp. 663–666). Portland, OR, USA.

Wang, B., Wang, L., & Qadir, T. (2011). Prosodic realization of focus in six languages/dialects in China. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 144–147). Hong Kong, China.

Wang, F. (2004). *Language contact and language comparison: the case of Bai* (Doctoral dissertation). City University of Hong Kong, Hong Kong.

Wang, F. (2005). On the genetic position of the Bai language. *Cahiers de Linguistique Asie Orientale*, *34*(1), 101–127.

Wang, W. S., & Li, K.-P. (1967). Tone 3 in pekinese. *Journal of Speech, Language, and Hearing Research*, *10*(3), 629–636.

Weinreich, U. (1974). *Languages in contact: Findings and problems*. The Hague: Mouton.

Wells, W. H. (1986). An experimental approach to the interpretation of focus in spoken English. *Intonation in Discourse*, *53*, 75.

Wiersma, G. C. (2005). *A study of the Bai (Minjia) language along historical lines* (Doctoral dissertation). University of California, Berkeley.

Wonnacott, E., & Watson, D. G. (2008). Acoustic emphasis in four year olds. *Cognition*, *107*(3), 1093–1101.

Wú, J. (吴积才), & Zhāng, F. (张茀). (1988). Dàlǐ fāngyán jiǎnkuàng jí yīnxì [A brief introduction of Dali Mandarin and its sound pattern, 大理方言简况及

音系]. *Yùxī Shīzhuān Xuébào [Journal of Yuxi Teachers' College,《玉溪师专学报》], 06*, 61–82.

Wu, W. L., & Chung, L. (2011). Post-focus compression in English-Cantonese bilingual speakers. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 148–151). Hong Kong, China.

Wu, W. L., & Xu, Y. (2010). Prosodic focus in Hong Kong Cantonese without post-focus compression. In *Proceedings of Speech Prosody 2010* (pp. 1–4). Chicago, IL, USA.

Xú, L. (徐琳) (Ed.). (2008). *Dàlǐ cóngshū: Báiyǔ piān [Dali series: Bai language 大理丛书: 白语篇]*. Kunming: Yúnnán Mínzú chūbǎn shè [Yunnan nationality publishing house 云南民族出版社].

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*(1), 55–105.

Xu, Y., Chen, S.-W., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review*, *29*(1), 131–147.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319–337.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197.

Yang, A. (2017). *The acquisition of prosodic focus-marking in Mandarin Chinese and Seoul Korean-speaking children* (Doctoral dissertation). Utrecht University.

Yang, A., & Chen, A. (2017). The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children. *First Language*. First Published October 11, 2017. https://doi.org/10.1177/0142723717733920

Yang, A., & Chen, A. (2014). Prosodic focus marking in child and adult Mandarin Chinese. In C. Gussenhoven, Y. Chen, & D. Dediu (Eds.), *Proceedings of*

the 4th International Symposium on Tonal Aspects of Language (pp. 54–58). Nijmegen.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Zerbian, S. (2007). Investigating prosodic focus marking in Northern Sotho. In K. Hartmann, E. Aboh, & M. Zimmermann (Eds.), *Focus strategies: Evidence from African languages* (pp. 55–79). Berlin: Mouton de Gruyter.

Zerbian, S. (2013). Prosodic marking of narrow focus across varieties of South African English. *English World-Wide*, *34*(1), 26–47.

Zhang, J. (2010). Issues in the analysis of Chinese tone. *Language and Linguistics Compass*, *4*(12), 1137–1153.

Zhang, J. (2014). Tones, tonal phonology, and tone sandhi. In J. C.-T. Huang, A. Y.-H. Li, & A. Simpson (Eds.), *The handbook of Chinese linguistics* (pp. 443–464). Malden, MA: John Wiley & Sons.

Zhāng, X. (张霞). (2012). HànBái shuāngyǔ jiàoyù jíqí jiàocái jiànshè [Bai-Chinese bilingual education and teaching material construction, 汉白双语教育及其教材建设]. *Dàlǐ Xuéyuàn xuébào[Journal of Dali University, 大理学院学报]*, *11*(7), 64–67.

Zhào, Yǎnsūn (赵衍荪), & Xú, L. (徐琳). (1996). *Bái-hàn cídiǎn [Bai-Chinese dictionary,《白汉词典》]*. Chengdu: Sìchuān mínzú chūbǎn shè [Sichuan Minzu Press, 四川民族出版社].

Zhào, Yànzhēn(赵燕珍). (2009). *Zhàozhuān Báiyǔ cānkǎo yǔfǎ [Zhaozhuan Bai reference grammar,《赵庄白语参考语法》]* (Doctoral dissertation). Minzu University of China.

Zhōu, F. (周福林), & Duàn, C. (段成荣) (2005). Wǒguó liúshǒu értóng zhuàngkuàng yánjiū [A study on children left behind 我国留守儿童状况研究]. *Rénkǒu Yánjiū [Population Research 人口研究]*, *29*(01), 29–36.

Zubizarreta, M. L., & Nava, E. (2011). Encoding discourse-based meaning: Prosody vs. syntax. Implications for second language acquisition. *Lingua*, *121*(4), 652–669.

APPENDICES

Appendix A: Parental questionnaire (Chinese)

云南喜洲镇白语韵律实验配套问卷 (家庭语言使用问卷) 20 题（包括表格）
家长姓名或代码_____ 孩子姓名或代码_____

Part A 个人背景情况

A1. 您的年龄.....岁

A2. 您的职业.....

A3. 您的最高学历.....

A4. 您的民族.....

A6N. 您的性别

□ 男

□ 女

A7N. 您和被调查人的关系是？

□ 父母与子女

□ 祖父母与孙子女

□ 如您的情况不属上述选项，请另作说明.....

Part B 语言能力 （可多选）

B2. 您的母语是什么？

□ 白语

□ 汉语普通话

□ 汉语方言

□ 如您的情况不属上述选项，请另作说明.....

B1. 您会说哪些语言？

□ 白语

□ 汉语普通话

□ 汉语方言

□ 如您的情况不属上述选项，请另作说明.....

Part C 语言学习途径

C1. 您的白语是怎样学会的？

□ 从小和抚养人学会的

（从小就会说，不知道是和谁学的）

□ 不是从小和抚养人学会的，是.....岁开始，

在.....（地方），

和.....（人）学的

□ 不会白语

□ 如您的情况不属上述选项，请另作说明.....

C2. 您的普通话是怎样学会的？

□ 从小和抚养人学会的

□不是从小和抚养人学会的，　　是.....岁开始，

在.....（地方），

和.....（人）学的

□ 不会普通话

□ 如您的情况不属上述选项，请另作说明.....

C3. 您的方言是怎样学会的？

□ 从小和抚养人学会的

□不是从小和抚养人学会的， 是.....岁开始，

在.....（地方），

和.....（人）学的

□ 不会方言

□ 如您的情况不属上述选项，请另作说明.....

C4. 您的.....语言怎样学会的？

□ 从小和抚养人学会的

□不是从小和抚养人学会的， 是.....岁开始，

在.....（地方），

和.....（人）学的

□ 如您的情况不属上述选项，请另作说明.....

Part D 语言环境

Part E 语言使用情况 （可多选）只要在您认同的方框里打勾即可

类别	序号	题目	普通话	汉语方言	白语	如您的情况不属上述选项，请另作说明
Part D 语言环境	D1N	家里常看的电视节目使用的语言？（可多选）				
	D5	家庭环境中，哪种语言占主要地位？				
Part E 语言使用情况	E6	与配偶常说哪种语言？（可多选）				
	E7	与子女/孙子女常说哪种语言？				
	E8	与家庭其他成员（除了父母、配偶和子女外）常说哪种语言				
	E9	与朋友常说哪种语言				

Part G 语言使用自我估算

G1. 您能对您使用白语的比例进行估算吗？ _____(例如 10%，20%.....100%)

G2. 您能对孩子在各个时段主要使用语言的情况加以估算吗？只要在您认同的方框里打勾即可

	时段	普通话	汉语方言	白语	如您的情况不属上述选项，请另作说明
上课期间	上学前 06:00 - 08:00				
	上课 08:00 - 12:00				
	中午休息 12:00 - 14:00				
	上课 14:00 - 17:00				
	放学后-回家 17:00 - 18:00				
	回家后 18:00 - 22:00				
周末休息	早 06:00 - 12:00				
	中 12:00 - 18:00				
	晚 18:00 - 22:00				

寒暑假	早 06:00 - 12:00
	中 12:00 - 18:00
	晚 18:00 - 22:00

Part H 孩子学前教育情况

H1. 孩子上过幼儿园

□ 小班

□ 中班

□ 大班

□ 孩子没上过幼儿园

□如您的情况不属上述选项，请另作说明……

H2. 孩子上过学前班

□ 1 年

□ 2 年

□ 孩子没上过学前班

□如您的情况不属上述选项，请另作说明……

Appendix B: Parental questionnaire (English)

Parental Questionnaire

Participants' name or code_____ Child's name or code_____

20 Items (tables included)

Part A Personal backgrounds

A1. Age .....

A2. Profession.....

A3. Education level.....

A4. Ethic group.....

A6N. Sex

- ☐ Male
- ☐ Female
- ☐ Others.....

A7N. What is your relationship with the child?

- ☐ Parents and children
- ☐ Grandparents and grandchildren
- ☐ Others.....

Part B Language competence (Multiple choice)

B2. What is your mother tongue?

- ☐ Bai language
- ☐ Putonghua
- ☐ Dali Mandarin

- ☐ Other languages.....

B1. What language/languages can you speak?

- ☐ Bai language
- ☐ Putonghua
- ☐ Dali Mandarin
- ☐ Other languages.....

Part C Approaches of language learning/acquisition

C1. How did you learn your Bai language?

- ☐ From my caregiver(s) when I was a kid.
- ☐When I was .....years old, in..... (place/location), from..... (people)
- ☐ I can't speak Bai language
- ☐ Other.....

C2. How did you learn your Putonghua?

- ☐ From my caregiver(s) when I was a kid.
- ☐When I was .....years old, in..... (place/location), from..... (people)
- ☐ I can't speak Putonghua
- ☐ Other.....

C3. How did you learn your Dali Mandarin?

- ☐ From my caregiver(s) when I was a kid.
- ☐When I was .....years old, in..... (place/location), from.....(people)
- ☐ I can't speak Dali Mandarin
- ☐ Other.....

C4. How did you learn your ..... (language)?

- ☐ From my supporter since I was a kid.
- ☐ When I was .....years old, in.....(place/location), from.....(people)
- ☐ I can't speak Bai language
- ☐ Other.....

D. Language environnement (Multiple choice)

E. Language use

Category	NO.	Item	Putonghua	Dali Mandarin	Bai	Other
D. Language environment	D1N	What language is used in the TV programs which your family mostly watches?				
	D5	What language is the major language used at home (in your family)?				
E. Language use	E6	What language or languages do you mostly use when talking to your partner? (Multiple choice)				
	E7	What language or languages do you mostly use when talking to your child/children or grandchild/grandchildren?				

E8	What language or languages do you mostly use when talking to your other family members (excluded your partner and children).	
E9	What language or languages do you mostly use when talking to your friends?	
E9	What language or languages do you mostly use when talking to your friends?	

Part G Self-estimation on language use

G1. Could you please make a self-estimation on the use of Bai language? _____(For example, 10%, 20%.....100%)

G2. Could you please make estimation about your children's language use? You could mark the cell which you think the language is mostly used by your children.

	Time period	Putonghua	Dali Mandarin	Bai	Other languages
School-days	Before school 06:00 - 08:00				
	At school 08:00 - 12:00				
	Lunch break 12:00 - 14:00				
	At school 14:00 - 17:00				
	After school – before coming home 17:00 - 18:00				
	After school 18:00 - 22:00				

Weekends	Morning 06:00 - 12:00
	Afternoon 12:00 - 18:00
	Evening 18:00 - 22:00
Summer and winter holidays	Morning 06:00 - 12:00
	Afternoon 12:00 - 18:00
	Evening 18:00 - 22:00

Thank you very much for your participation!

Appendix C: Transcription of experimental materials in Bai

Note: Here only one target sentence in five different focus conditions is illustrated. All the subjects (n = 4), verbs (n = 6) and objects (n = 4) can be substituted for eliciting other target sentences. All the subjects, verbs and objects can be found in Chapter 2 Table 1.

NF-i:

Sentence 1:

/nɔ^{33} xa^{33}/!

you look !

'Look!'

Sentence 2:

/tsɯ^{33}/.

Tree.

'Tree.'

Sentence 3:

/tse^{55} tsɯ^{35} kɯ^{21} xuo^{44} ŋɣ̩^{33} lɯ^{44} tɕa^{55} tɕa^{55} le^{21}/.

and sell things POSS shelf one (quantifier)

'and a shelf.'

Sentence 4:

/sa^{42} ʑy^{35} a^{55} tɯ^{21} tsɯ^{33} lɯ^{33} ui^{33} kɯ^{21} tsɯ^{33}/.

seems someone at here sell tree.

'It seems like that someone sells the tree.'

Sentence 5:

/a^{55} tɯ^{21} kɯ^{21} tsɯ^{33}/ ?

who one sell tree ?

'who sells the tree?'

NF-m:

Sentence 1:

/nɔ^{33} xa^{33} /!

you look !

‘Look !’

Sentence 2:

/thɔ^{55} lɔ^{55}/.

Rabbit.

‘Rabbit.’

Sentence 3:

/tse^{55} tsɯ^{35} tsɯ^{33}/.

and tree one (quantifier).

‘And tree’.

Sentence 4:

/sa^{42} ʑy^{35} thɔ^{55} lɔ^{55} tɯ^{21} ȵo^{44} tɔ^{42} tsɯ^{33}/.

seems rabbit one want do tree.

‘It seems like that the bear does something to the tree.’

Sentence 5:

/thɔ^{55} lɔ^{55} tɯ^{21} tsɿ^{55} mɯ^{55} tɔ^{42} tsɯ^{33}/?

Rabbit one how do tree?

‘What does the bear do to the tree?’

<u>NF-f:</u>

Sentence 1:

/nɔ^{33} xa^{33}/!

You look!

‘Look!’

Sentence 2:

/ɕo^{42}/.

Bear.

‘Bear.’

Sentence 3:

/ɕo^{42} tɯ^{21} tsɯ^{31} ke^{33} kɯ^{21} xuo^{44} ŋɣ^{33} lɯ^{44} tɕa^{55}tɕa^{55} le^{21}/.

bear one at be sell things POSS shelf one.

'The bear stands beside a shelf'.

Sentence 4:

/sa^{42}ʑy^{35} ɕo^{42} tɯ^{21} tsɯ^{33} lɯ^{33} ui^{33} kɯ^{21} xuo^{44} ŋɣ^{33}/.

seems bear one at here sell things

'It seems like that the bear sells something.'

Sentence 5:

/ɕo^{42} tɯ^{21} kɯ^{21} xa^{55} le^{21}/?

bear one sell what one?

'What does the bear sell?'

<u>BF:</u>

Sentence 1:

/nɔ^{33} xa^{33}/!

You look!

'Look!'

Sentence 2:

/ŋɔ^{33} xa^{55} ʑa^{33} ȵi^{55} xa^{33} tɯ^{33} tuo^{33}/.

I what all look not.

'I can't see anything clearly'.

Sentence 3:

/lɯ^{33} lɔ^{44} xua^{55} tso^{35} lɔ^{44} xua^{55} tɯ^{33} xa^{55} le^{21}/?

you POSS picture one POSS draw doing what one?

'What has been depicted in your picture?'

<u>CF-m:</u>

Sentence 1:

/nɔ^{33} xa^{33}/!

You look!

‘Look!’

Sentence 2:

/khua^{33}/.

Dog.

‘Dog.’

Sentence 3:

/tse^{55} tsɯ^{35} tsɯ^{33}/.

and tree one.

‘And tree.’

Sentence 4:

/sa^{42} ʑy^{35} khua^{33} tɯ^{21} ȵo^{44} tɔ^{42} tsɯ^{33}/.

seems dog one want do tree.

‘It seems like that the bear does something to the tree.’

Sentence 5:

/ŋɔ^{33} tsɔ^{42} khua^{33} tɯ^{21} ma^{35} tsɯ^{33}/.

I say dog one wipe tree.

‘I guess the bear wipes the tree.’

Appendix D: Overview of participants (N = 51)

Speaker groups	Age groups	Participant ID	Age	AoA	Gender	No.of Total trials	Usable responses (duration & pitch analysis)	% (duration & pitch analysis)[36]	Usable responses (pitch analysis)	% (pitch analysis)
Bai-Mandarin early bilingual children	6-7	g1_p03	7	7	M	80	8	10%	7	9%
Bai-Mandarin early bilingual children	6-7	g1_p04	7	7	F	80	28	35%	25	31%
Bai-Mandarin early bilingual children	6-7	g1_p05	7	7	M	80	51	64%	46	58%
Bai-Mandarin early bilingual children	6-7	g1_p06	6	6	F	80	66	83%	53	66%
Bai-Mandarin early bilingual children	6-7	g1_p07	7	7	M	80	53	66%	52	65%
Bai-Mandarin early bilingual children	6-7	g1_p08	7	7	F	80	45	56%	45	56%
Bai-Mandarin early bilingual children	6-7	g1_p10	7	7	F	80	46	58%	41	51%
Bai-Mandarin early bilingual children	6-7	g1_p16	7	7	M	80	46	58%	42	53%
Bai-Mandarin early bilingual children	9-10	g3_p01	9	6	M	80	70	88%	69	86%
Bai-Mandarin early bilingual children	9-10	g3_p03	9	6	M	80	67	84%	61	76%
Bai-Mandarin early bilingual children	9-10	g3_p04	9	6	F	80	64	80%	62	78%
Bai-Mandarin early bilingual children	9-10	g3_p07	9	6	F	80	64	80%	63	79%
Bai-Mandarin early bilingual children	9-10	g3_p12	9	6	F	80	59	74%	51	64%
Bai-Mandarin early bilingual children	9-10	g3_p13	9	6	M	80	53	66%	52	65%
Bai-Mandarin early bilingual children	9-10	g3_p14	9	6	F	80	75	94%	67	84%
Bai-Mandarin early bilingual children	9-10	g3_p15	10	7	M	80	62	78%	61	76%

[36] "% (duration & pitch analysis)" indicates the proportion of the number of usable responses for duration and pitch analysis to the total number of trials produced by each speaker. Similarly, "% (pitch analysis)" indicates the proportion of the number of usable responses for pitch analysis to the total number of trials produced by each speaker.

Speaker groups	Age groups	Participant ID	Age	AoA	Gender	No.of Total trials	Usable responses (duration & pitch analysis)	% (duration & pitch analysis)[36]	Usable responses (pitch analysis)	% (pitch analysis)
Bai-Mandarin early bilingual children	12-13	g6_p01	12	6	M	80	68	85%	62	78%
Bai-Mandarin early bilingual children	12-13	g6_p02	12	6	F	80	65	81%	59	74%
Bai-Mandarin early bilingual children	12-13	g6_p04	13	7	M	80	62	78%	58	73%
Bai-Mandarin early bilingual children	12-13	g6_p05	12	6	F	80	56	70%	53	66%
Bai-Mandarin early bilingual children	12-13	g6_p06	13	7	M	80	53	66%	48	60%
Bai-Mandarin early bilingual children	12-13	g6_p08	12	6	F	80	76	95%	75	94%
Bai-Mandarin early bilingual children	12-13	g6_p09	13	7	F	80	69	86%	55	69%
Bai-Mandarin early bilingual children	12-13	g6_p12	13	7	M	80	67	84%	60	75%
Bai-Mandarin early bilingual children	12-13	g6_p13	12	6	M	80	60	75%	54	68%
Bai-Mandarin early bilingual adult	adult	ta_p02	44	6	F	80	72	90%	67	84%
Bai-Mandarin early bilingual adult	adult	ta_p03	45[37]	6	F	80	70	88%	69	86%
Bai-Mandarin early bilingual adult	adult	ta_p04	43	6	F	80	74	93%	73	91%
Bai-Mandarin early bilingual adult	adult	ta_p05	46	6	M	80	73	91%	65	81%
Bai-Mandarin early bilingual adult	adult	ta_p06	29	6	F	80	76	95%	66	83%
Bai-Mandarin early bilingual adult	adult	ta_p07	51	6	M	80	67	84%	52	65%
(near-) monolingual speaker of Dali Mandarin	adult	xgaa_p03	28	from birth	F	80	69	86%	67	84%
(near-) monolingual speaker of Dali Mandarin	adult	xgaa_p04	33	from birth	M	80	66	83%	66	83%
(near-) monolingual	adult	xgaa_p05	54	from birth	M	80	75	94%	74	93%

[37] This participant declined to report her age. Her age is an estimation.

Speaker groups	Age groups	Participant ID	Age	AoA	Gender	No.of Total trials	Usable responses (duration & pitch analysis)	% (duration & pitch analysis)[36]	Usable responses (pitch analysis)	% (pitch analysis)
speaker of Dali Mandarin (near-) monolingual speaker of Dali Mandarin (near-) monolingual speaker of Dali Mandarin (near-) monolingual speaker of Dali Mandarin (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai (near-) monolingual speaker of Bai	adult	xg aa_p06	28	from birth	M	80	78	98%	68	85%
	adult	xgaa_p07	28	from birth	F	80	77	96%	73	91%
	adult	xgaa_p08	33	from birth	F	80	75	94%	67	84%
	adult	p01	23	from birth	M	60	46	77%	38	63%
	adult	p02	24	from birth	M	60	54	90%	52	87%
	adult	p03	25	from birth	M	60	41	68%	35	58%
	adult	p04	22	from birth	F	60	44	73%	36	60%
	adult	p05	24	from birth	M	60	55	92%	37	62%
	adult	p06	24	from birth	M	60	42	70%	37	62%
	adult	p07	46	from birth	F	60	53	88%	46	77%
	adult	p08	37	from birth	F	60	46	77%	33	55%
	adult	p09	45	from birth	F	60	42	70%	41	68%
	adult	p10	39	from birth	M	60	52	87%	44	73%
	adult	p11	23	from birth	F	60	54	90%	43	72%
	adult	p12	40	from birth	F	60	50	83%	47	78%
	adult	p13	26	from birth	F	60	60	100%	42	70%
	adult	p14	24	from birth	F	60	58	97%	51	85%

SAMENVATTING

Er is een grote groep kinderen die opgroeien met meer dan één taal, zowel in de wereld als in China. De kennis van de taalontwikkeling van tweetalige kinderen is echter nog beperkt, vooral op het gebied van prosodie. Een centrale vraag in onderzoek naar tweedetaalverwerving van vroeg tweetalige kinderen is of de tweedetaalverwerving van vroeg tweetalige kinderen lijkt op de eerstetaalverwerving van eentalige kinderen. Dit proefschrift gaat over de prosodische ontwikkeling in de tweede taal van vroeg tweetalige kinderen in het algemeen en in het bijzonder over de verwerving van een belangrijke functie van prosodie, namelijk om nieuwe informatie aan te geven of om focus aan te geven in een zin. Dit onderzoek bestudeert het ontwikkelingspad en uiteindelijke niveau van prosodische focusmarkering in het Mandarijn van Bai (L1) - Mandarijnse (L2) vroeg tweetaligen.

Bai (Sino-Tibetaanse taalfamilie, Bai genus) is een toontaal die wordt gesproken door de Bai etnische groep (Bai volk, Báizú 白族) met name in de Dali Bai Autonome Prefectuur (*Dàlǐ báizú zìzhìzhōu* 大理白族自治州), Yunnan (*Yúnnán* 云南), in het zuidwesten van China. Mandarijn (Sino-Tibetaanse taalfamilie, Chinese genus) is een toontaal die wordt gesproken door het Han/Chinese (Hànzú 汉族) volk in het hele gebied van China. Standaard Mandarijn, ook wel Putonghua (Pǔtōnghuà 普通话) genoemd, is de standaardvariëteit van Mandarijn en de officiële nationale taal van de Volksrepubliek China. Standaard Mandarijn is het leerdoel van de tweedetaalverwerving van vroeg Bai-Mandarijnse tweetaligen. In de Dali Bai Autonome Prefectuur waar Bai-Mandarijnse vroeg tweetaligen opgroeien, wordt naast Bai en Standaard Mandarijn ook veel Dali Mandarijn (*Dàlǐ fāngyán* 大理方言), een regionale variëteit van het Mandarijn, gesproken. Dali Mandarijn wordt vooral door het Han volk in het Dali gebied gesproken, met name in de stad Dali, en het klanksysteem ervan is vergelijkbaar met Standaard Mandarijn. Alle Bai sprekers zijn tweetaligen die Dali Mandarijn en/of Standaard Mandarijn als hun tweede taal

kunnen spreken. Er zijn echter ook sprekers van het Dali Mandarijn die weinig kennis van het Bai hebben en in het stedelijk gebied van de stad Dali (*Xiàguān* 下关) wonen. Hoewel de promotie van het Standaard Mandarijn in China begonnen is vanaf de jaren 50, waren Bai en Mandarijn al eeuwenlang in langdurig en intensief contact in Dali. Bai-Mandarijnse tweetaligen groeien in een tweetalige taalomgeving op waarin Bai vooral thuis met familieleden gebruikt wordt, Standaard Mandarijn vooral op school voor onderwijsdoeleinden en passief ontvangen door massa media, en Dali Mandarijn wordt als omgevingstaal gesproken door het Han volk in de omgeving van de tweetaligen.

Centraal in het huidige onderzoek staat de vraag hoe vroeg tweetaligen prosodische focusmarkering verwerven in hun tweede taal. Om precies te zijn, ons doel was om inzicht te krijgen in twee belangrijke thema's: (1) Laat tweedetaalverwerving van vroeg tweetalige kinderen een zelfde snelheid en route van verwerving zien als de eerstetaalverwerving van eentalige kinderen? (2) Hoe beïnvloedt de L1- en L2-input de verwerving van prosodische focusmarkering door vroeg tweetalige kinderen? Om hierachter te komen bestudeerden we, door gebruik te maken van een experimentele benadering, het gebruik van prosodie om nauwe focus en focustypes die in constituentgrootte en –contrastrerendheid verschillen te coderen, in verschillende talen en taalvariëteiten die door verschillende groepen sprekers gesproken worden.

Om dit proefschrift in het juiste perspectief te plaatsen, gaven we in Hoofdstuk 1 een overzicht van het veld van prosodische focusmarkering door het begrip focus, de taalkundige realisatie van focus en het gebruik van prosodische kenmerken om focus te markeren in Aziatische toontalen, met een focus op Standaard Mandarijn, te introduceren. Een review van voorgaande onderzoeken over de verwerving van prosodische focusmarkering door tweetaligen en een introductie van de sociolinguïstische situatie en taalkundige (prosodische) kenmerken van L1-Bai en L2-input (Semi-Standaard Mandarijn en Dali Mandarijn) van Bai-Mandarijnse vroeg tweetaligen werden gepresenteerd in Hoofdstuk 1. We eindigden Hoofdstuk 1 met het formuleren van onderzoeksvragen, hypotheses en de opbouw van het proefschrift.

Hoofdstuk 2 onderzocht prosodische focusmarkering in Zuidelijk Bai met als doel om eerstehands informatie van de L1 van Bai-Mandarijnse vroeg tweetaligen te leveren. We hebben gevonden dat (semi-) eentalige sprekers van Bai de duur van de focalische constituent verlengen om nauwe focus van post-focus te onderscheiden. Ze variëren echter geen toonhoogte-gerelateerde kenmerken om nauwe focus te coderen. Daarnaast gebruiken ze geen toonhoogte-gerelateerde kenmerken en ook geen duur om focustypes (nauwe focus, contrastrerende focus, en brede focus) van elkaar te onderscheiden. Het is bekend dat eentalige sprekers van Standaard Mandarijn nauwe focus prosodisch markeren door de toonhoogtevariatie uit te breiden en de duur van de focalische constituent te verlengen, hoewel ze alleen duur variëren om nauwe focus van brede focus te onderscheiden. In vergelijking met sprekers van Standaard Mandarijn gebruiken sprekers van Bai prosodie in mindere mate. Onze resultaten impliceren dat Bai geen invloed lijkt te ondergaan van Standaard Mandarijn in prosodische focusmarkering, ondanks het feit dat Bai in een langdurig en intensief contact met Mandarijn is geweest en lexicale en syntactische invloed van Mandarijn laat zien.

Hoofdstuk 3 onderzocht prosodische focusmarkering in Dali Mandarijn en had als doel om informatie te geven over kenmerken van L2-input (d.w.z., gesproken talen of taalvariëteiten in de omgeving van kinderen) in de tweedetaalverwerving van Bai-Mandarijnse tweetalige kinderen. We vonden dat (semi-) eentalige sprekers van Dali Mandarijn de duur van de focalische constituent verlengen in vergelijking met zijn tegenhanger in de post-focus en pre-focus conditie, maar geen toonhoogte-gerelateerde kenmerken variëren om nauwe focus te coderen. Daarnaast variëren (semi-) eentalige sprekers van Dali Mandarijn noch toonhoogte-gerelateerde kenmerken noch duur om focustypes die verschillen in constituentgrootte en –contrastrerendheid te onderscheiden, vergelijkbaar met (semi-) eentalige sprekers van Bai (Hoofdstuk 2), maar anders dan eentalige sprekers van Standaard Mandarijn. Onze resultaten laten zien dat Dali Mandarijn, dat in langdurig en intensief contact met Bai is geweest, prosodie op eenzelfde manier gebruikt om focus te coderen als

Bai. Anders dan in veel regionale variëteiten van Mandarijn en Standaard Mandarijn, wordt in Dali Mandarijn focus prosodisch gecodeerd door middel van duur.

Hoofdstuk 4 concentreerde zich op prosodische focusmarkering in Mandarijn gesproken door Bai-Mandarijnse vroeg-tweetalige kinderen in drie leeftijdsgroepen: zes- tot zevenjarigen, negen- tot tienjarigen en twaalf- tot dertienjarigen. Dit Hoofdstuk had als doel om het ontwikkelingspad van prosodische focusmarkering in het Mandarijn van Bai-Mandarijnse vroeg-tweetalige kinderen te onthullen. We vonden dat Bai-Mandarijnse vroeg-tweetalige kinderen het gebruik van duur om nauwe focus te coderen verwerven als ze tussen de zes en zeven jaar oud zijn. Daarnaast variëren ze noch duur noch toonhoogte-gerelateerde prosodische kenmerken om nauwe focus van brede focus te onderscheiden, onafhankelijk van toon. Tot slot, geen van de leeftijdsgroepen gebruiken duur of toonhoogte-gerelateerde prosodische kenmerken om contrastieve focus van niet-contrastieve focus te onderscheiden, onafhankelijk van toon. Onze resultaten laten dus zien dat Bai-Mandarijnse vroeg tweetalige kinderen het gebruik van duur vroeger verwerven dan het gebruik van toonhoogte-gerelateerde prosodische kenmerken voor focus-markering doeleinden in hun Mandarijn. We denken dat het relatief makkelijker is om het gebruik van duur te beheersen dan toonhoogte als een prosodisch kenmerk voor focus-markering doeleinden, omdat er een beperkte akoestische ruimte beschikbaar is voor focus-gerelateerde manipulatie in toonhoogte voor leerders van toontalen in een vroeg stadium van hun taalontwikkeling. Verder, omdat de L1-Bai van Bai-Mandarijnse vroeg tweetalige kinderen duur gebruikt om nauwe focus te coderen, vermoeden wij dat het gebruik van duur als een prosodisch kenmerk om nauwe focus te coderen in de L1 van tweetaligen de verwerving van het gebruik van duur om nauwe focus te coderen makkelijker maakt in de L2 van tweetaligen. Onze resultaten vonden echter geen bewijs dat Bai-Mandarijnse vroeg tweetalige kinderen het gebruik van prosodie om focustypes te onderscheiden eerder verwerven dan om nauwe focus van non-focus te onderscheiden. Omdat de L1-Bai van Bai-Mandarijnse vroeg tweetalige kinderen noch toonhoogte noch duur gebruikt om focustypes die in constituentgrootte verschillen te coderen, vermoeden wij dat de onsuccesvolle verwerving van prosodische kenmerken om focustypes te coderen in

de L2 van de tweetalige kinderen verklaard kan worden door negatieve L1 transfer. Verder zou de niet-Standaard Mandarijn-achtige input ook een belangrijke rol kunnen spelen.

Hoofdstuk 5 bestudeerde prosodische focusmarkering in Mandarijn gesproken door de Bai-Mandarijnse vroeg tweetalige volwassenen die de leraren waren van de Bai-Mandarijnse vroeg tweetalige kinderen. Door deze groep van tweetaligen te onderzoeken, probeerden we het uiteindelijke niveau van prosodische focusmarkering in Bai-Mandarijnse vroeg tweetaligen vast te stellen en informatie te geven over de L2-input van tweetalige kinderen. We vonden dat Bai-Mandarijnse vroeg tweetalige volwassenen de duur van de focalische constituent verlengen om nauwe focus van pre- en post-focus te onderscheiden, vergelijkbaar met de productie van eentalige Mandarijn-sprekende volwassenen. Daarnaast vergroten ze het toonhoogtebereik en verhogen ze de maximum toonhoogte van de focalische constituent om nauwe focus van post-focus te onderscheiden, terwijl eentalige Mandarijn-sprekende volwassenen toonhoogte-gerelateerde kenmerken en duur gebruiken om nauwe focus van zowel pre- als post-focus te onderscheiden. Anders dan eentalige Mandarijn-sprekende volwassenen, gebruiken onze Bai-Mandarijnse vroeg tweetalige volwassenen echter geen toonhoogte-gerelateerde kenmerken noch duur om focustypes te coderen die in constituentgrootte en –contrastrerendheid verschillen. We hebben vastgesteld dat Bai-Mandarijnse vroeg tweetalige volwassenen uiteindelijk de vaardigheid behalen om toonhoogte en duur te variëren om focus te coderen, maar niet op een volledig Standaard Mandarijn-achtige manier. Het lijkt erop dat deze overeenkomst tussen de L1 en de L2 van tweetaligen het leren van het duurkenmerk in Mandarijn makkelijker kan maken. Onze Bai-Mandarijnse vroeg tweetaligen hebben echter laten zien dat ze een opmerkelijke vaardigheid hebben om het toonhoogtebereik uit te breiden en de maximum toonhoogte van de focalische constituent te verhogen om nauwe focus van post-focus te onderscheiden in hun Mandarijn. We denken dat de intensieve training in Standaard Mandarijn de vaardigheid van Bai-Mandarijnse vroeg tweetaligen kan

verklaren om toonhoogte-gerelateerd kenmerken te variëren om nauwe focus te coderen.

Hoofdstuk 6 gaf een samenvatting van de belangrijkste resultaten die in Hoofdstuk 2, 3, 4, en 5, beschreven zijn, besprak de details van de overeenkomsten en verschillen in de snelheid en route van verwerving tussen de L2 van vroeg tweetalige kinderen en de L1 van eentalige kinderen en de invloed van L1 en L2-input, bespraken de implicaties van de bevindingen voor het veld van tweetaligheid, en gaf suggesties voor toekomstig onderzoek.

Wat betreft de overeenkomsten en verschillen in het ontwikkelingspad van prosodische focusmarkering tussen de L2 van Bai-Mandarijnse vroeg tweetalige kinderen en de L1 van Mandarijn-sprekende eentalige kinderen, zijn de tweetalige kinderen vergelijkbaar met de eentalige kinderen in hun relatief vroegere beheersing van duur dan toonhoogte om nauwe focus te coderen in hun L2-Mandarijn verwerving. In tegenstelling tot Mandarijn-sprekende kinderen die het gebruik van prosodie om focustypes te onderscheiden eerder verwerven dan het onderscheid tussen nauwe focus en non-focus, variëren Bai-Mandarijnse tweetalige kinderen noch duur noch toonhoogte om focustypes te onderscheiden op zes- tot zevenjarige leeftijd. Op dezelfde leeftijd (zes tot zeven jaar oud), beheersten de tweetalige kinderen echter het gebruik van duur om nauwe focus van non-focus (d.w.z., pre- en post-focus) in hun Mandarijn te onderscheiden. Wat betreft de ontwikkelingssnelheid vonden we verschillen in de verwervingssnelheid tussen de L2 van vroeg tweetaligen en de L1 van eentaligen. Bai-Mandarijnse vroeg tweetalige kinderen hebben niet een vergelijkbare competentie ontwikkeld op twaalf- tot dertienjarige leeftijd (d.w.z., na vijf jaar van officieel Standaard Mandarijn onderwijs) als vier tot vijf jaar oude eentalige Mandarijn-sprekende kinderen met een vergelijkbare periode van blootstelling aan Standaard Mandarijn. Dat laat zien dat tweetaligen een relatief langzamere verwervingssnelheid hebben. In tegenstelling tot de beweringen van eerdere studies op de gebieden van vocabulaire en grammaticale verwerving van de L2 van vroeg tweetaligen dat de route en snelheid van de verwerving van de L2 van vroeg tweetaligen vergelijkbaar zijn met

de verwerving van de L1 van eentalige kinderen, maar in overeenkomst met de beweringen van eerdere studies in de gebieden van de verwerving van segmentele fonologie in de L2 van vroeg tweetaligen, vonden we dat tweetaligen niet vergelijkbaar zijn met eentaligen, en relatief langzamer zijn in de ontwikkeling van hun L2.

Er zijn twee aannemelijke oorzaken voor de relatief langzame verwervingssnelheid in de L2 van Bai-Mandarijnse vroeg tweetalige kinderen. Ten eerste, het verschil tussen Bai (L1) en Mandarijn (L2) in prosodische focusmarkering kan het verschil verklaren in verwervingssnelheid. Het gebruik van toonhoogte voor lexicale doeleinden is aanwezig in zowel de L1 (Bai) als de L2 (Mandarijn) van Bai-Mandarijnse tweetaligen, terwijl het gebruik van toonhoogte voor prosodische focusmarkering op zinsniveau alleen in hun L2 (Mandarijn) aanwezig is. Daardoor kan de verwerving van taalspecifiek gebruik van prosodie om focus te coderen in hun L2 een uitdagende taak zijn voor Bai-Mandarijnse vroeg tweetaligen. Ten tweede, de niet-Standaard Mandarijn-achtige input zou een rol kunnen spelen in de relatief langzame verwervingssnelheid van L2-Mandarijn van Bai-Mandarijnse vroeg tweetaligen. Eentalige Mandarijn-sprekende kinderen die in Beijing leven en opgroeien zijn niet alleen blootgesteld aan Beijing Mandarijn maar ook aan Standaard Mandarijn in alledaagse communicatie en via massa media. Bai-Mandarijnse vroeg tweetaligen daarentegen groeien op in een gebied dat relatief geïsoleerd is van Standaard Mandarijn in het zuidwesten van China. Hoewel ze passieve toegang hebben tot Standaard Mandarijn via massa media, zijn ze vooral blootgesteld aan niet-Standaard variëteiten van het Mandarijn (d.w.z., Semi-Standaard Mandarijn en Dali Mandarijn) in alledaagse communicatie. We lieten zien dat geen van deze variëteiten van het Mandarijn focus markeren op een Standaard Mandarijn-achtige manier. Eerdere studies van de competentie van vroeg tweetalige volwassenen in prosodische focusmarkering speculeerden alleen dat de onsuccesvolle verwerving van prosodische focusmarkering in de L2 van vroeg tweetaligen verklaard zou kunnen worden door de "niet-Standaard-achtige input". Dit proefschrift bestudeerde echter de verschillende vormen van L2-input in de

taalomgeving van vroeg tweetalige kinderen. Door dat te doen, leverde dit proefschrift voor de eerste keer direct empirisch bewijs dat de kwaliteit van L2-input ook de verwervingssnelheid in de ontwikkeling van de L2 van vroeg tweetalige kinderen kan beïnvloeden. De niet-Standaard input zou op zijn beurt bijdragen aan het niet-moedertaal-achtige gebruik van toonhoogte in Mandarijn door Bai-Mandarijnse vroeg tweetalige volwassenen met een hoge vaardigheid in Mandarijn.

In conclusie, door Bai-Mandarijnse vroeg tweetalige kinderen van zes tot dertien jaar oud en tweetalige volwassenen te onderzoeken, hebben we het ontwikkelingspad en het uiteindelijke niveau van prosodische focusmarkering in het Mandarijn van Bai-Mandarijnse vroeg tweetaligen vastgesteld. We hebben zowel verschillen als overeenkomsten gevonden in de route en snelheid van verwerving van prosodische focusmarkering tussen de L2 van vroeg tweetalige kinderen en de L1 van eentalige kinderen. Daarnaast hebben we gevonden dat Bai-Mandarijnse vroeg tweetalige volwassenen erg vaardig zijn in het gebruik van duur en toonhoogte-gerelateerde prosodische kenmerken om focus te coderen in het Mandarijn, maar dat ze niet volledig Standaard Mandarijn-achtig zijn. Onze resultaten laten zien dat de invloed van de L1 (d.w.z., positieve en negatieve transfer) duidelijk aanwezig is in de ontwikkeling van de L2 van tweetaligen, wat uitgebreid geobserveerd is voor tweetalige taalverwerving in verschillende taalkundige gebieden. Belangrijker, niet-Standaard L2 input kan ook de route en snelheid van verwerving in de L2 ontwikkeling van vroeg tweetaligen beïnvloeden.

CURRICULUM VITAE

Zenghui Liu was born on April 15th 1987 in Yunnan, China. In 2005, she started to study Teaching Chinese as a Second Language at Guizhou Normal University, China. From 2007 to 2008, she was selected as an exchange student to study at Daegu University, South Korea. In 2009, she obtained a bachelor's degree with a double major in Teaching Chinese as a Second Language and English Education. After finishing her bachelor, she was awarded a three-year scholarship and recommended to Yunnan University to continue her MA study. In 2012, she got a master's degree in Chinese Philology and Linguistics at Yunnan University. In 2014, her MA thesis titled *The Study of Zhu Dexi's Grammar Research* was awarded as the Outstanding MA Thesis of Yunnan Province. In 2012, she was award a four-year scholarship from the Chinese Scholarship Council and started her PhD research at the Utrecht Institute of Linguistics OTS, Utrecht University. During her PhD she presented her work at several international workshops and conferences. This dissertation is the result of work that Zenghui carried out between 2012 and 2017 as a PhD candidate at Utrecht University.