1 Introduction

Mother: Daddy hai6 me1 jan4 aa3? 'What's Daddy's nationality?'
Child: Ing1gok3jan4 'English person.'
Mother: Jing1gok3jan4 'English person.'
Child: Ing1gok3jan4 'English person.'
Maa1mi4 hai6 zung1gok3jan4 'Mummy is Chinese.'
Mother: Timmy hai6 me1 jan4 aa3? 'What about Timmy?'
Child: Bilingual! (Timmy 2;00;14)

1.1 Introduction

Talking to a young bilingual child can be both entertaining and eye-opening. Even at the tender age of two, the bilingual child is capable of expressing complex ideas, having two languages at his disposal as seen in the above exchange between Timmy and his mother (the first author). Timmy refers to his father as *ing1gok3jan4* 'English person', his mother as *zung1gok3jan4* 'Chinese person' and he surprises everyone, not least his mother, by referring to himself as *bilingual*.[2] Apart from raising deep issues of awareness of identity, this exchange epitomises an important phenomenon typical of a bilingual child, namely that he is in contact with two languages on a daily basis.

What is it like being a bilingual child? How do children cope with learning two languages simultaneously in the first years of life? Many children, like those of cross-cultural marriages, grow up in families where more than one language is spoken on a regular basis. Their parents may each speak a different language natively, thus exposing these children to two languages from birth. The principal protagonists of this book are three siblings born in such a family where the mother is a native speaker of Cantonese (the first author) and the father of British English (the second author). As the parents of the children, we have the advantage of observing their language development on a daily basis, making a first-hand eye-witness account possible. As linguists specializing in language acquisition and language contact respectively, we have followed our bilingual children's emerging language from their first forms of vocalization to

mastery of complex syntax. Following their linguistic odyssey in these golden years of language acquisition, one can only wonder at the inexorable process of acquisition gradually unfolding before our eyes and ears. With the help of modern technology, their language development over time has been captured and recorded in the form of audio and video-recordings. The corpus containing these three siblings' transcriptions of longitudinal recordings from the age of 1;03 and 3;06 forms the primary empirical basis for the present study. Known as the *Hong Kong Bilingual Child Language Corpus*, it documents the longitudinal development of a total of six bilingual children growing up in Hong Kong and is available via the *Child Language Data Exchange System* (CHILDES).[3] At the time of writing, it was the largest multimedia bilingual corpus and also 'the largest corpus of linked video data on child language development available in any language (in CHILDES)' (Brian MacWhinney p.c.).[4] In addition, we have kept our own diary of our observations of their progress. Taken together, we have assembled not only an endless repertoire of anecdotes, but also a wealth of data which provide compelling evidence for a set of propositions about these children's bilingual development, including the following:

- While the two languages are differentiated from early on, there is strong evidence for syntactic transfer and interaction between the two linguistic systems developing in the mind of the bilingual child.
- There are principles determining the direction of transfer and mechanisms which account for how it takes place: these include *language dominance*, *developmental asynchrony* and *input ambiguity*. The cross-linguistic influence evidenced in the bilingual development is bidirectional, going primarily from the *dominant language* to the *non-dominant language* but in certain domains also from the non-dominant to the dominant language. We shall refer to the non-dominant language as weaker language interchangeably throughout the book.
- The developmental patterns in bilingual individuals parallel and reflect prominent features of contact varieties, such as *Singapore Colloquial English*, spoken by a community of adult bilingual speakers at the societal level. This comparison in turn sheds light on processes and mechanisms of language contact at large.

This book presents a series of case studies in early bilingual development involving a so far largely unstudied and divergent pair of languages, Cantonese and English, focusing on some features which shed light on the nature and processes of bilingual development. This is the first systematic longitudinal study of Cantonese-English bilingualism in childhood covering the children's language development in the first three years, extending to five and beyond in some cases. Just as bringing a wider range of languages into consideration changes our view of what is possible in human languages, so it promises to change our view of what is possible in language development. The bulk of previous

research on bilingual development in the period from zero to three years has been focused on European languages. The languages in both classic and recent longitudinal studies include Indo-European language pairs such as English – German (Leopold 1939–1949, Döpke 1992), English – Dutch (De Houwer 1990), English – Norwegian (Lanza 2004), English – Spanish (Deuchar & Quay 2000), French – German (Ronjat 1913, Meisel 1990, 1994), French – Serbian (Pavlovitch 1920) and German – Italian (Taeschner 1983).[5] Of the thirty-odd longitudinal studies listed by Hoffman (1991), from Ronjat (1913) to De Houwer (1990), all but four involved Indo-European language pairs; notable exceptions include Smith (1931; 1935),[6] the first case study involving Chinese and English, and Vihman (1985) on English-Estonian bilingual development. Chang-Smith (2005) compares the development of a Mandarin-English bilingual child with that of a monolingual Mandarin-speaking child in a study of nominal expressions in Mandarin. These studies have been revealing in many respects, but in terms of global linguistic diversity, they have investigated only a tiny fraction of the possible language combinations a child might be faced with. The ways in which a typologically divergent language pair such as Cantonese and English differ open up possibilities for interaction which would not exist with other language pairs. The numerous fundamental contrasts between the two languages provide potential for cross-linguistic influence and transfer in various grammatical domains of acquisition which form the focus of our case study. The study of bilingual development involving a Chinese language will contribute to diversification of language pairs in the study of childhood bilingualism, providing a new window for viewing developmental processes and pathways and enriching both the theoretical investigation and empirical coverage of early bilingual acquisition.

1.1.1 Practical and cognitive implications

The study of this particular language pair is also of growing practical importance, since the number of bilingual families with children speaking English and Cantonese, Mandarin or another Chinese language is on the rise in the twenty-first century. They represent a significant population of children around the world who share similar bilingual experiences as our children in this study. At a more general level, the study of Cantonese-English bilingual development can be seen as an instantiation of bilingual development in a broader sense: what is observed here should be to some extent generalizable to other cases of bilingual development.

Another category of children who are drawing increasing attention from the international academic community is that of adopted children whose language development before and after adoption has become an intriguing domain of inquiry. Recent years have seen the rising number of international adoptions

around the world, with China being the number one source of children adopted into the United States. Many of the adoptees from China into American families leave their home at infancy before age two or three and have to acquire a 'second first language' (Pollock, Price & Fulmer 2003; Roberts et al. 2005). In monolingual English-speaking homes, it is likely that these children's first language, Chinese, will gradually be lost while English takes the place of Chinese as their first language (Nicoladis & Grabois 2002). For those adopted into homes with Chinese spoken regularly and English in the community, some form of bilingualism is likely to develop, with both Chinese and English acquired together. Questions arise as to whether these constitute cases of bilingual or child second language acquisition (as discussed in chapter 2). Similarly, preschool immigrant children who move from Chinese-speaking communities to an English-speaking country or vice versa will have the opportunity to develop childhood bilingualism. Li and Lee (2002) investigate the development of Cantonese in British-born Chinese-English bilinguals and report delayed and stagnated development of Cantonese due to incomplete learning of their L1 Cantonese and influence of English, a dominant language in the environment. The present study may shed light on language acquisition by these populations given that a Chinese language and English are involved across these acquisition contexts. A recent study of childhood bilingualism in Korean immigrant children in America by Shin (2004) shows that the children 'follow similar but delayed patterns of first language acquisition of Korean and second language acquisition of English' (Shin 2004: 12), while bidirectional influence is found, with L1 Korean influencing the development of L2 English which in turn influences the development of Korean.

At the general level, childhood bilingualism offers many cognitive advantages for the developing bilingual child. From the perspective of cognitive development and language processing, Bialystok (2001) examines various linguistic and cognitive consequences of developing two languages in childhood, discussing the potential contribution of childhood bilingualism in illuminating the nature of linguistic knowledge, organization of cognitive processes and the functional structure of the brain. Among the issues covered are developmental issues in language acquisition, metalinguistic awareness, literacy and problem solving. She explores and highlights the complexities and intricacies that make the empirical study of bilingual development so challenging, arguing that bilingual children are different from monolinguals in the way they acquire language and concluding that 'The vast majority of cognitive differences were advantageous to the bilingual children' (Bialystok 2001: 232). Her views also echo Grosjean's (1989) insight that 'bilinguals are not two monolinguals in one'. According to Grosjean's holistic view of bilingualism, the bilingual is not the sum of two complete or incomplete monolinguals but an integrated whole with a unique linguistic profile.

1.2 Research questions

Acquiring two languages in childhood holds endless fascination for lay people and specialists alike. As the leading American structuralist Leonard Bloomfield (1933: 29) remarked, the acquisition of language 'is doubtless the greatest intellectual feat any one of us is ever required to perform'. If a child's acquisition of a language is a miracle, then acquiring two at the same time is doubly so. Given that our children have become fluent speakers of two languages in the space of a few years, one cannot help but wonder how they accomplish this feat. Language acquisition by children has been compared to natural and effortless activities like walking and recognizing faces which in fact involve complex mental processes and mechanisms. The naturalness and inevitability of the outcome is compared to the perception of solid objects and attention to line and angle by Chomsky (1965: 59). The ability of the child to acquire language is what Pinker (1994) calls the 'language instinct': knowledge of language is not acquired as a result of teaching, but is to a large extent attributable to the human innate capacity for language acquisition. The field of first language acquisition has been far from unanimous regarding what exactly is attributed to nature vs. nurture, which will continue to be one of the central themes of debate in the years to come. We remain open as to how to characterize this language instinct. While Chomsky and Pinker see the language instinct as specific to the language faculty, an alternative possibility is that articulated by Bates and MacWhinney (1989: 10):

> The human capacity for language could be both innate and species-specific, and yet involve no mechanisms that evolved specifically and uniquely for language itself. Language could be viewed as a new machine constructed entirely out of old parts.

This alternative view espouses explanations that are not domain-specific, but encompass general cognition, processing and neuro-cognitive functions as new research findings continue to challenge much of our received wisdom. We do not venture to take a definitive position on which aspects of knowledge of language are derived from domain-specific innate Universal Grammar and which from domain-general mechanisms. The issue of how to characterize the nature of linguistic knowledge will be further discussed in chapter 2.

In the present context of bilingual development, we shall refer to the *bilingual instinct*, the language instinct given full expression in the simultaneous acquisition of two languages by children. It is simply human, and totally natural, for the bilingual child to acquire both languages in response to the dual input in the environment. Compared with acquiring one language in monolingual contexts, the acquisition of two languages in bilingual or multilingual contexts poses even more challenges to the child on many grounds, beginning with the fact that the

quantity of input in each language is necessarily reduced by around one half (cf. Paradis & Genesee 1996). We shall show that the processes involved in the simultaneous construction of two grammars in the child's mind are inherently different from that of constructing one grammar only. Bilingual children often take a different path from the monolingual counterparts to reach the target, as is clearly instantiated in the case studies discussed in the following chapters.

The questions we address in this work include the following:

- How does bilingual development differ from acquisition of the same two languages by monolingual children?
- Do the two languages develop independently or do they interact systematically? Is there evidence for transfer or cross-linguistic influence? What factors determine the direction of transfer?
- What can the linguistic features of bilingual children's developing languages reveal about more general processes in language acquisition and language contact?

In studying our own children we are following a time-honoured tradition beginning with the classic studies of Ronjat (1913) and Leopold (1939–49). Ronjat (1913) inaugurated Grammont's principle, *une personne, une langue*, i.e. the one parent – one language approach in addressing the bilingual child.[7] Ronjat's longitudinal study of his own son Louis' development in French and German is generally considered the earliest bilingual study in the twentieth century (see Hoffmann 1991: 50–53). Werner Leopold, a professor of German with the combined passion of a father and a developmental psychologist, recorded the bilingual development of his daughters Hildegard and Karla in German and English, culminating in the monumental work *Speech Development of a Bilingual Child: A Linguist's Record* published in four volumes between 1939 and 1949 and containing over eight hundred pages of intense and close observation of bilingual development in early childhood. Without the help of a tape recorder, Leopold recorded his daughters' speech data in the form of a diary with extensive commentaries on specific linguistic features. Leopold's linguistic study of early bilingual development remains unparalleled in terms of the comprehensive coverage of the details of a child's simultaneous acquisition of two languages. Even today, Leopold is held in high esteem as one of the founding fathers of the study of bilingualism as well as of child language at large. Leopold felt that the study of child language would reveal much about general principles of language and language change: 'every pattern of grammar, every process of language shows up in child language in a nascent state, in coarser, more tangible shapes, compressed into a much shorter time and therefore more accessible to observation'. Leopold's legacy will always remain a source of inspiration and serve as an important reference for case studies in bilingual development.

Building on the foundations established by our predecessors and inspired by their vision, we have conducted a longitudinal study of our own children using the recording techniques and apparatus available and feasible at the time. The case-study approach continues to be fruitful: contemporary studies in the field of bilingual development include De Houwer's (1990) study of a Dutch-English bilingual child, Lanza's (2004) case study of two bilingual children from Norwegian-American families and Deuchar and Quay's (2000) study of Deuchar's English-Spanish bilingual daughter, all of which fall squarely in this tradition of longitudinal case studies.

We are convinced that the advantages of studying one's own children outweigh the disadvantages. The advantages include:

- privileged access to the children throughout, and beyond, the period of study;
- first-hand knowledge of the children's environment and experiences;
- the unique dual status of linguists and parents (doubled in the case of both parents being linguists);

Among these advantages, it is only thanks to the diary data that we are able to document the emergence of English prenominal relative clauses, which are scarcely found in the regular longitudinal recordings. We shall see, in the case study of relative clauses (chapter 6), how shared knowledge between parent and child is a prerequisite for the felicitous use of this construction. We also take responsibility for ethical issues such as privacy (and trust that our children will understand). The drawbacks include:

- potential for subjectivity (for example, in selection and transcription of diary data);
- enhanced potential for rich interpretation of the data (for example, in attributing more advanced knowledge to the children than they have actually demonstrated);
- the Observer's Paradox, whereby the very presence of the observer changes the situation being observed (Labov 1972).

A poignant example of the Observer's Paradox is the case of the parent-researcher going away to record in the diary what the child has just said, thereby interrupting the conversation and changing the course of events. There is inevitably a trade-off here since one needs to record the utterances while they are still fresh in the mind, within seconds or minutes of the utterances being produced.

1.3 The ecology of bilingual development

The social context in which acquisition takes place to a large extent determines the input to the child and the outcome. This is especially important in

bilingual and multilingual contexts where the nature and quantity of input in each language, and the prevailing attitudes to each, all contribute to determining processes and outcomes of development. For example, the prevalence of code-mixing in children's language depends on both its occurrence in the adult input, and adult attitudes to it (Lanza 2004).

The notion of ecology, applied by Mufwene (2001) primarily to the evolution of languages in contact, is equally applicable to the development of individual bilingualism.[8] Ecology here begins as a metaphor from biology: the environment in which languages are spoken determines the course of development of languages, much as habitats determine the evolution and fate of species in competition with each other. This point is undoubtedly applicable even to monolingual contexts:[9] for example, social factors such as prestige may determine the selection of variant forms leading to sound change (Nettle 1999). It is still more salient and important, however, in determining the outcomes of language contact situations such as those discussed by Mufwene (2001), and the cases of bilingual development at issue here. This is because the range of variants from which linguistic options may be selected (the 'feature pool' in Mufwene's terms, see chapter 2) is so much wider compared to monolingual contexts. In the case of a bilingual environment, the feature pool is in principle doubled, or even (to the extent that code-mixing and intermediate options exist) more than doubled. In the case of creoles:

> The ethnographic ecology . . . affected the role of the external structural ecology toward more, or less, influence, as it determined the particular conditions under which it was possible for a language to influence the restructuring of the target language. (Mufwene 2001: 161)

Ecology in this sense refers to the social environment in which a language is spoken. The external ecology of a language encompasses all other languages with which its speakers come into contact, the number of speakers of each language and their social status. Mufwene (2001: 21–24) further extends the notion of ecology to internal factors affecting the evolution of language. Within languages, 'Linguistic features in a system also constitute part of the ecology for one another' (Mufwene 2001: 22). Internal ecology in this sense is again analogous to a related concept in biology where ecology can be taken as internal to a species. For example, dialectal variation and co-existent systems within a language all impact the evolutionary trajectory of a language.

In the context of bilingual development, internal ecology involves the competition between, and selection of, variants available in language systems (with some variants being made available through transfer from another language system). Consider, for example, the acquisition of *wh*-interrogatives as discussed in chapter 4. Between ages two and three, the child has two forms of *wh*-question competing with each other (Yip & Matthews 2000a: 199):

(i) The target *wh*-movement represented in the input, as in *What is this for?*
(ii) The *wh*-in-situ form transferred from Cantonese, as in *This is for what?*
With the external ecology (input from adult speakers of English) supporting option (i), our bilingual children eventually select option (i) over (ii). Given a community of bilingual speakers using option (ii), however, the child might select the *wh*-in-situ form (ii) instead, or allow both forms to co-exist. Just such a community of bilingual speakers exists in the case of contact varieties such as Singapore Colloquial English, as discussed in chapter 4.

1.4 The Hong Kong speech community

The children of the present study were born and raised in Hong Kong. Cantonese is the community language of Hong Kong spoken by around 90% of its residents.[10] According to figures given in the entry [Chinese, Yue] in the *Ethnologue* (Gordon 2005: 331), native speakers of Yue dialects (the dialect group to which Cantonese belongs) in all countries amount to some 55 million, ranking 16th in the top 100 languages by population. A former British colony for over 150 years, Hong Kong continued to recognize English as an official language, along with Cantonese and Mandarin, after the handover of sovereignty to China in 1997. The official language policy of Hong Kong is for its citizens to be 'biliterate and trilingual', speaking Cantonese, Putonghua and English and being literate in both English and standard written Chinese.

Among Hong Kong people who are ethnic Chinese, Cantonese is the lingua franca. In the Hong Kong Chinese community, many children like our own grow up in an extended family situation (as they do in Singapore, cf. Gupta 1994). Since the relatives speak primarily Cantonese, the children's everyday environment provides more input in this community language than in English.

It should be noted that Cantonese is essentially a spoken language. To the extent that Cantonese is written down at all, it is heavily affected by standard written Chinese, which is based on Mandarin. A tradition of vernacular literature exists using Chinese characters to represent Cantonese as it is spoken, but such writing has low status (Snow 2004). Many colloquial morphemes in Cantonese do not have a corresponding character in the written language, though attempts have been made to standardize usage and fill the gaps (Cheung & Bauer 2002). Reference works on Cantonese grammar include Cheung (1972), Matthews and Yip (1994) and Yip and Matthews (2000b, 2001).

In this book, we are solely concerned with the acquisition of the spoken language, leaving aside the acquisition of literacy in bilingual development.[11] Many aspects of Cantonese and Mandarin child language development, including both spoken language and literacy, are covered in Li et al. (2006).

Like individual speakers, bilingual communities including Hong Kong are best characterized along a continuum of bilingualism. In a multilingual

community such as Hong Kong, it is common to speak of bilingualism as a matter of degree. Parents, for example, may discuss children's bilingualism using code-mixing:

(1) Keoi5 go3 zai2 zan1hai6 hou2 *bilingual* gaa3
she CL son really very bilingual SFP
'Her son is really very bilingual.'

A commercial radio station advertised its bilingualism in similar terms:

(2) Disc jockey 1: Next, we have bilingual news.
Disc jockey 2: Hai6 aa3, hou2 *bilingual* aa3
is PRT very bilingual SFP
'Yes, very bilingual.'

Compared to Singapore, for example, the use of English in the Hong Kong speech community is relatively restricted: it is used widely in secondary and higher education, the higher courts and international companies, but rarely on the street, or even in markets or shopping malls, outside typical tourist haunts. Much more widely used than pure English is code-mixing, in which English terms (such as *bilingual* in the above examples) appear within a Cantonese sentence structure (Li 1996; B. Chan 1998; 2003). With a long history of contact between English and Cantonese, code-mixing has been a ubiquitous phenomenon in educated Hong Kong speech (Li & Lee 2004).

In Hong Kong, as in Singapore, Chinese dialects other than Cantonese form part of the picture, typically being spoken by older relatives as well as recent immigrants from mainland China. In the case of our own children, the Chaozhou dialect is spoken by their grandmother and relatives of her generation and above; the children had some passive knowledge of it, but produced it rarely, usually for jocular effect. For example, inserting a Chaozhou phrase produces a trilingual utterance:

(3) Gong2 Ciu4zau1 waa2 is *puah lok k'u*, fall down. [laughs]
'Speaking in Chiu Chow, *puah lok k'u* means "fall down".'
(Timmy 2;02;10)

For the most part, influence of other southern Chinese dialects on English will be similar to that deriving from Cantonese: all Chinese dialects exhibit certain broad typological traits such as *wh*-in-situ, null arguments and prenominal relative clauses, all of which will be central to our analyses of the bilingual children's syntactic development. In certain domains, however, the roles of different dialects can and should be differentiated. Min dialects of Chinese such as Hokkien and Chaozhou (known as *Teochew* in Singapore) are particularly divergent, and may account for specific features of Singapore Colloquial English (SCE). For example, questions of the form 'X or not?' produced