J. Child Lang. **43** (2016), 553–580. © Cambridge University Press 2016
doi:10.1017/S0305000915000562

Chinese L1 children's English L2 verb morphology over time: individual variation in long-term outcomes

JOHANNE PARADIS*, YASEMIN TULPAR AND ANTTI ARPPE

University of Alberta

(*Received* 29 *October* 2014 – *Revised* 21 *July* 2015 – *Accepted* 21 *September* 2015 – *First published online* 26 *February* 2016)

ABSTRACT

This study examined accuracy in production and grammaticality judgements of verb morphology by eighteen Chinese-speaking children learning English as a second language (L2) followed longitudinally from four to six years of exposure to English, and who began to learn English at age 4;2. Children's growth in accuracy with verb morphology reached a plateau by six years, where 11/18 children did not display native-speaker levels of accuracy for one or more morphemes. Variation in children's accuracy with verb morphology was predicted by their English vocabulary size and verbal short-term memories primarily, and quality and quantity of English input at home secondarily. This study shows that even very young L2 learners might not all catch up to native speakers in this time frame and that non-age factors play a role in determining individual variation in child L2 learners' long-term outcomes with English morphology.

INTRODUCTION

The common wisdom concerning second language (L2) learning is "the younger, the better", with child L2 learners being expected to quickly and successfully catch up to their native-speaker peers. Research on age of acquisition onset (AoA) largely supports the common wisdom when it comes to comparing the L2 proficiency of individuals who began to learn a L2 in adulthood vs. childhood (e.g. DeKeyser, 2012). However, less is known about whether learning a L2 at different AoAs within the

[*] We would like to thank the families and children for their time and effort in participating in this research. We would also like to thank Ruiting Jia, Katryna Lysay, Emily Yiu, and Tatiana Zdorenko for their assistance with data collection. This research was funded by the Social Sciences and Humanities Research Council of Canada (Standard Research Grant #410-2010-0123 to Paradis), for which we are grateful. Address for correspondence: Johanne Paradis, University of Alberta – Linguistics, 4-57 Assiniboia Hall, Edmonton, Alberta T6G 2E7, Canada. e-mail: johanne.paradis@ualberta.ca

childhood years makes a difference in how long it takes for children to become identical to native speakers of that language, and if all of them do so. Contrary to what is commonly expected, some studies have shown that adults who began to learn the societal language as a L2 in childhood do not always possess L2 grammar and pronunciation equivalent to native speakers (e.g. Abrahamsson & Hyltenstam, 2009). Also contrary to the maxim of "the younger, the better", early AoA is no guarantee of developing and maintaining native-like proficiency in a language, since heritage language speakers can experience attrition or incomplete acquisition (Montrul, 2008). Moreover, researchers have found that language input and experiential factors, first language (L1) background, and language-learning aptitude also shape children's L2 development, possibly more so than AoA (e.g. Jia & Fuse, 2007; Paradis, 2011; Unsworth, Argyri, Cornips, Hulk, Sorace & Tsimpli, 2014). Longitudinal long-term outcome studies have rarely been conducted with child L2 learners, and the impact of non-AoA factors on child L2 learners has mainly been studied during the early stages of L2 acquisition. Accordingly, this study sought to determine if English L2 children, all of whom had AoAs in early childhood, would reach native-speaker levels of accuracy with English verb morphology after long-term exposure to English as a L2 in English-medium schools in an English majority-language city. Our secondary goal was to determine if non-AoA individual difference factors would predict variation in children's L2 morphological abilities at this late stage in L2 acquisition, and in so doing, understand if these factors influence whether or not L2 children converge on native-speaker accuracy with verb morphology in the elementary school years.

Age effects in L2 acquisition

Lenneberg (1967) proposed that maturational, neurological changes around adolescence cause the offset of a critical period for language acquisition such that a language is rarely learned to native-speaker proficiency after this age. Since Lenneberg's proposal, there has been a great deal of research and debate about how AoA impacts L2 acquisition. Researchers currently debate whether AoA effects in L2 attainment are caused by maturational (internal) or external factors, whether AoA impacts grammatical subdomains differentially, and whether there is a specific 'cut-off' age or whether AoA effects are continuous across the lifespan (for reviews, see DeKeyser, 2012; Muñoz & Singleton, 2011; Tsimpli, 2014). The research most relevant to the present study concerns the relationship between AoAs in childhood and long-term L2 outcomes.

Several studies with adults have found that non-native ultimate attainment can occur in individuals who began to learn a L2 in early childhood

(Abrahamsson & Hyltenstam, 2009; Flege, Munro & MacKay, 1995; Flege, Yeni-Komishan & Liu, 1999; Hakuta, Bialystok & Wiley, 2003; McDonald, 2000; Weber-Fox & Neville, 1999). First, Abrahamsson and Hyltenstam (2009) examined both the grammatical and pronunciation abilities of 195 Spanish first language (L1)–Swedish L2 speakers who had been living in Sweden for at least ten years, but had started to learn Swedish at different ages. They found differences in grammar and pronunciation between Swedish native speakers and Swedish L2 learners even for those L2 learners who began to learn Swedish at age 5;0 or younger, although discrepancies with native-speaker abilities increased along with AoA. Similar to Abrahamsson and Hyltenstam (2009), McDonald (2000) found non-native attainment in grammatical knowledge for Vietnamese L1–English L2 adults, including some who began to learn English at or before age 5;0. Regarding pronunciation, Flege *et al*. (1995) and Flege *et al*. (1999) showed that perceivable foreign accents increased continuously with increasing AoA in English L2 speakers with long-term residence in North America, but importantly, perceivable foreign accents were found in some individuals with AoAs < 5;0. However, Flege *et al*. (1999) found discrepancies with native speakers in English L2 grammatical abilities only for individuals with AoAs > 5;0. Hakuta *et al*. (2003) examined self-reported general proficiency in English in US census data from millions of respondents whose L1s were either Spanish or Chinese. The graphical data and analyses show that the respondents' English proficiency was indisputably related to the age when they began to learn English, and this decline with age began in the early childhood period. Finally, Weber-Fox and Neville (1999) report a series of studies with Chinese-L1–English-L2 adults where differences between native speakers and L2 speakers were found in grammatical test scores and neuro-processing as a function of increasing AoA, beginning with AoAs of 4;0–6;0. Taken together, this group of studies demonstrates that AoAs in early childhood do not necessarily predict uniform convergence with native-speaker outcomes in L2 grammar and pronunciation. Instead, they suggest that the likelihood of native-like attainment decreases gradually and continuously along with increasing AoA.

Most studies of AoA in L2 acquisition have a developmental retrospective design – that is, they include adult participants with various AoAs – and just a few prospective studies on AoA with L2 children have been conducted to date (Armon-Lotem, Walters & Gagarina, 2011; Jia & Fuse, 2007; Meisel, 2008, 2009; Unsworth, 2013; Unsworth *et al*., 2014). Meisel (2008, 2009) observed non-native or L2 acquisition patterns in German-L1–French-L2 children with AoAs between 3;0 and 4;0, and proposed that this age range might mark the end of an early sensitive period for L2 morphological acquisition; however, this proposal was based on age-related differences in

error patterns in the early stages of French L2 acquisition, not in long-term attainment in French. Unsworth (2013) and Unsworth *et al.* (2014) did not find that different AoAs influenced bilingual children's abilities with grammatical gender in Dutch and Greek as L2s, except some differences emerged between simultaneous-from-birth bilinguals and early L2 learners in Greek. In contrast, Armon-Lotem *et al.* (2011) found negative correlations between AoA and L2 grammatical abilities in German and Hebrew by Russian L1 children. However, these studies by Unsworth, Armon-Lotem, and their colleagues included some children with low exposure to the L2, and thus were not exclusively examining AoA effects in late-stage L2 acquisition. Regarding longitudinal research, Jia and Fuse (2007) found that Chinese-L1 children with AoAs in early childhood had higher levels of accuracy with English L2 verb morphology in spontaneous speech over a five-year period than those with AoAs in late childhood/adolescence. However, the early AoA advantage only emerged for two of six grammatical constructions examined, and the small sample size (N = 10) and heterogeneous AoA spread (5;0–16;0) complicate the conclusions that can be drawn from this study. In sum, methodological issues and conflicting findings indicate more prospective developmental research with child L2 learners is needed to determine when, and under what conditions, they diverge from, or converge with, native speakers in their grammatical abilities. While the studies with adults cited above indicate that divergence can be the long-term outcome for some early AoA L2 speakers, we do not know at what point in development such divergence emerges.

Non-age factors influencing child L2 acquisition

Marinova-Todd, Marshall, and Snow (2000) argue that non-age factors can shape L2 development and outcomes as much or more than AoA. The developmental retrospective studies discussed in the previous section indicate that non-native outcomes for L2 speakers are probabilistic in that there is variation among early L2 learners in their ultimate attainment. This raises the issue of what non-AoA factors contribute to determining L2 acquisition outcomes. Sources of individual differences in L2 abilities can be either environmental, i.e. external to the child, or they can reflect inherent linguistic and cognitive abilities internal to the child. Regarding external factors, individual differences in input quantity, i.e. amount or length of L2 exposure, have been found to predict differences in children's L2 grammatical abilities (Armon-Lotem, Joffe, Abutbul-Oz, Altman & Walters, 2014; Armon-Lotem *et al.*, 2011; Blom & Paradis, 2015; Blom, Paradis & Sorenson Duncan, 2012; Bohman, Bedore, Peña, Mendez-Perez & Gillam, 2010; Chondrogianni & Marinis, 2011; Hoff, Welsh, Place &

Ribot, 2014; Marinis & Chrondrogianni, 2010; Paradis, 2011; Unsworth, 2013; Unsworth *et al*., 2014). Quality of linguistic input and experience also play a role in child L2 acquisition. Higher family socioeconomic status, greater richness of the L2 environment (e.g. frequency and diversity of reading, media use, organized activities, and playing with friends in the L2), greater parental fluency in the L2, and having older siblings in school are associated with stronger L2 grammatical abilities (Armon-Lotem *et al*., 2011; Bohman *et al*., 2010; Chrondrogianni & Marinis, 2011; V. M. Gathercole, 2007; Hoff *et al*., 2014; Jia & Aaronson, 2003; Jia & Fuse, 2007; Paradis, 2011). Importantly, Jia and Fuse (2007), Unsworth (2013), and Unsworth *et al*. (2014) found that input and experiential factors explained more variance in children's L2 grammatical abilities than AoA in their studies. Similarly, in their study with adult L2 speakers, Flege *et al*. (1999) found that years of education in the United States more strongly predicted English grammatical outcomes than AoA.

In addition to child-external input factors, child-internal factors also influence rate of L2 acquisition. First, research has indicated that children with Chinese L1s (Mandarin and Cantonese) are slower to acquire English L2 verb morphology within the first three years of exposure than children with other L1 backgrounds (Blom & Paradis, 2013, 2015; Blom *et al*., 2012; Paradis, 2011). These researchers argued that the more protracted English L2 development of Chinese L1 speakers is likely because their languages lack grammatical tense and subject–verb agreement marking and are typologically isolating; therefore, these L2 learners are not experiencing positive transfer from the L1 to the L2. McDonald's (2000) retrospective developmental study also found an L1 effect because Spanish-L1–English-L2 speakers outperformed Vietnamese-L1–English-L2 speakers in their knowledge of English verb inflection, even when both groups' AoAs were in early childhood; furthermore, Spanish-L1–English-L2 speakers performed more like native speakers. A second child-internal factor predicting L2 acquisition is verbal memory skills, a component of language learning aptitude (Dörnyei & Skehan, 2003). Researchers have found verbal memory skills to be predictive of L2 outcomes in instructed/foreign L2 learners (Harley & Hart, 1997; Masoura & S. Gathercole, 1999), and correlated with monolingual children's ability to detect errors with verb morphology (McDonald, 2008). Paradis (2011) found that English L2 children's verbal short-term memory was the strongest predictor of individual differences in their accuracy with English verb morphology in the first three years of learning English, even stronger than length of exposure to the L2. A third child-internal factor associated with L2 grammatical acquisition is vocabulary size. Research with both simultaneous bilingual and L2 children has revealed that size of vocabulary in a language is associated with grammatical development in

the same language (Conboy & Thal, 2006; Marchman, Martínez-Sussmann & Dale, 2004; Simon-Cereijido & Gutiérrez-Clellen, 2009). More specifically, there is an association between L2 vocabulary size and accuracy with L2 verb inflections such as third singular [-s] and past tense (Blom & Paradis, 2013; Blom *et al.*, 2012; Marinis & Chondrogianni, 2010).

In addition to child-level factors, language-level factors can also influence children's accuracy with verb morphology. Language-level factors refer to frequency and distributional properties of the input that all speakers/hearers would be exposed to, and thus are not sources of individual differences at the child level. For the present study, we considered language-level factors pertaining to the inflectional morphemes third singular [-s], regular past [-ed], and past irregular, such as the frequency of an inflected word (verb stem + affix or irregular past, *dug* or *ran*) in the input and allomorph type (third singular -s: [s], [z], [ɪz]; past regular: [t], [d], [ɪd]). Research has shown that the acquisition of English L2 grammatical morphology is sensitive to word frequency and allomorph type (Blom & Paradis, 2013; Blom *et al.*, 2012; Goldschneider & DeKeyser, 2001; Marinis & Chrondrogianni, 2010). Regarding frequency of inflected words, this means that L2 learners would be more accurate in using third singular [-s] or the past tense with a verb that appears more frequently in this inflectional form in the input than with another verb that appears less frequently. Regarding allomorph types, these are phonologically conditioned by the verb stem, but are also unevenly distributed in the input since there are more verb stems that take the voiced obstruent [z] or [d] than take the VC allomorphs [ɪz] or [ɪd]. For example, verb stems ending in either voiced consonants or vowels take [z], whereas, stems ending in sibilants take [ɪz], and the former comprises a larger set of verb stems in English (Blom & Paradis, 2013; Blom *et al.*, 2012). It is difficult to disentangle whether phonological or type frequency factors contribute to the later acquisition of the VC allomorphs (cf. Blom & Paradis, 2013), and so for the present study we did not explore this issue.

To date, most of the research examining the impact of non-age age factors in child L2 has either focused on early stages of L2 acquisition or has included both early- and late-stage L2 children in the study sample. Consequently, it is not well known to what extent these individual difference (external and internal) and language-level factors continue to influence L2 abilities at later stages of acquisition.

Present study: design and research questions

This study sought to determine if English L2 children with AoAs in early childhood would all reach native-speaker levels of accuracy with English verb morphology in production and with judgements of correct/incorrect

https://doi.org/10.1017/S0305000915000562 Published online by Cambridge University Press

use after long-term L2 exposure. Children's accuracy with verb morphology was examined over three years; a longitudinal design was chosen so that the shape of developmental trajectories could be examined. Regarding length of L2 exposure, we based our choice of time frame on previous research indicating that L2 children catch up to native speakers in their oral language abilities after approximately four to six years of exposure in preschool/school (Hakuta Goto Butler & Witt, 2000; Saunders & O'Brien, 2006). More specifically concerning morphology, Jia and Fuse's (2007) study of accuracy with English L2 grammatical morphology showed a plateau or asymptote in development after four to five years of exposure for most morphemes. Also, Marinis and Chondrogianni (2010) found that Turkish-L1–English-L2 children were close or equivalent to native-speaker accuracy with verb inflection by six years of exposure. Accordingly, for the present study, children were examined in their fourth, fifth, and sixth year of exposure to English in preschool/school.

The predictive role of environmental factors and the internal factors of verbal short-term memory and vocabulary size on children's long-term outcomes was examined in this study. Only Chinese L1 children were included in the study. This was because of their protracted acquisition of English morphology demonstrated in previous research (Blom *et al*., 2012; Jia & Fuse, 2007; Paradis, 2011). Therefore, L1 background was not manipulated as an individual-difference variable. Regarding language-level factors, the impact of word frequency and allomorph on children's accuracy with verb inflection in production was examined.

In brief, English L2 children with Chinese (Mandarin and Cantonese) L1 backgrounds were given tests of production and grammaticality judgements with verb morphology once a year for three years/rounds. These were standardized tests, normed with monolinguals. Monolinguals reach ceiling on these tests by age 6;0, and their scores remain stable and at ceiling as they get older (Rice & Wexler, 2001; Rice, Wexler & Hershberger, 1998). Therefore, assessing how native-like the performance of L2 children with four to six years of exposure to English is on these tests constitutes a fair comparison. Analyses of these longitudinal data focused on addressing the following questions:

1. Is there change across the three rounds in children's scores on the verb morphology tests? If so, does L2 learning appear to be growing or reaching a plateau?
2. Have the L2 children reached native-speaker levels of abilities with verb morphology by the final round?
3. What environmental and child-internal factors influence children's accuracy with L2 verb morphology? What language-level frequency factors influence use of L2 verb inflection?

METHOD

Participants

Child participants were recruited from Cantonese- and Mandarin-speaking families residing in Edmonton, Canada. There are close to 100 non-English languages spoken in Edmonton and 6·8% and 3·9% of the population report speaking Cantonese and Mandarin, respectively, at home (Statistics Canada, 2011). Parents had to be both foreign-born and native speakers of a Chinese language and L2 speakers of English. Children were either Canadian-born or foreign-born, but had to have started sustained and consistent exposure to English in a daycare, preschool, or school programme before age 6;0. While families varied in their use of English at home at the time of testing, as an inclusion criterion, all children had to have been spoken to exclusively or primarily in Chinese by their parents from birth until at least 3;0. Thus, there were no families who had bilingual language use at home starting from the child's birth, but instead the children were raised with primarily Chinese at home and English at daycare/preschool/school in their early years. In general, our sample could be characterized as having a high socioeconomic status background because the majority of the mothers had some post-secondary education. Mean maternal education in years was 14·6 (SD = 3·3). This is not unusual given Canada's point-based immigration system, where higher education levels increase the likelihood of acceptance for immigration.

Participants were chosen to form a cohort with respect to AoA, chronological age, and length of exposure from a larger database of children who had participated in previous studies. First, AoA had to be < 6;0 for inclusion, and in our sample the mean AoA was 4;2 (SD = 1;0, range = 1;7–5;8). For chronological age, we aimed for a mean of approximately 8½ years at Round 1, and recruited children whose age was no more than ±12 months of this mean. For length of exposure at Round 1, we aimed for a mean of approximately 4½ years, and recruited children whose length of exposure was no more than ±6 months of this mean. Children's actual chronological ages and length of exposure to English at each round are presented in Table 1.

In the sample, 10/18 children had Cantonese as their L1 and 8 had Mandarin as their L1. Both languages are typologically isolating and do not mark tense grammatically or have subject–verb agreement morphology (Lin, 2001; Matthews & Yip, 1994); furthermore, previous research has found that children from both these Chinese language backgrounds display more protracted acquisition of verb morphology in English than children whose L1s are typologically inflecting and mark tense and agreement grammatically (Blom *et al.*, 2012; Paradis, 2011). Nevertheless, we examined if there were any consistent between-group differences based on

https://doi.org/10.1017/S0305000915000562 Published online by Cambridge University Press

TABLE 1. *Participant characteristics*

	Round 1	Round 2	Round 3
Age	8;5 (0;11)	9;5 (0;11)	10;5 (0;11)
Length of exposure	4;3 (0;6)	5;3 (0;6)	6;4 (0;7)
English-use-at-home	.36 (0·27)	.37 (0·25)	.40 (0·25)
English richness	.63 (0·08)	.68 (0·09)	.65 (0·09)
CTOPP – non-word repetition	7·9 (1·6)	9·3 (1·2)	9·1 (2·4)
PPVT-IV	99 (16)	105 (16)	109 (19)

NOTES: Age is chronological age. Length of exposure means years and months of exposure to English in daycare/preschool/school. Proportion of English use in the home, spoken to the child by family members and spoken by the child to family members, is calculated between 0 and 1·0, with 1·0 as only English being used/spoken. Richness of the English environment is calculated between 0 and 1·0, with 1·0 as the richest possible English environment. CTOPP is Wagner *et al*. (1999), and the non-word repetition subtest is a measure of verbal short-term memory. These are standard scores, mean = 10, 1 SD range = 7–13. The PPVT-IV is Dunn and Dunn (2007) and measures receptive vocabulary size. These are standard scores, mean = 100, 1 SD range = 85–115.

L1 for our dependent variables in the present study. Non-parametric comparisons were used because of small Ns. Mann–Whitney *U* tests comparing the scores between Cantonese- and Mandarin-speaking children for all TEGI probes (see 'Materials and procedure'; the past tense probe was divided into regular and irregular scores) at all rounds were conducted, 24 comparisons in total. Results yielded 23/24 non-significant differences and one marginally significant difference, for third singular [-s] at Round 3 ($p = .059$). Based on this analysis, we judged that grouping the children together was justified.

Materials and procedure

Children were tested in their homes once a year, and parents were given a questionnaire during the home visits. The tests measured children's abilities with verb morphology, their verbal short-term memory, and their receptive vocabulary size. The parent questionnaire was used to obtain information on a child's quantity and quality of English input. The verb morphology constructions targeted in testing were: third person singular (3rd sing. -s), *he walks;* past regular, *he walked* and past irregular, *he ran*; BE auxiliary and copula, *they are walking, he is happy*; DO auxiliary, *does he walk every day?* What follows is a description of the tests and the questionnaire used to obtain our dependent and independent variables.

Test of Early Grammatical Impairment (TEGI). The TEGI (Rice & Wexler, 2001) was administered to the children, following the instructions in the Examiner's Manual. The TEGI includes production probes for the use of 3rd sing. [-s], past regular [-ed], past irregular, BE and DO (in

questions and statements). The TEGI also has grammaticality judgement (GJ) probes for correct use, incorrect use, or omission of verb morphemes.

For the 3rd sing. [-s] probe, children were shown pictures of professionals engaged in work activities and given prompts like *Here is a teacher. Tell me what a teacher does*. Expected answers included *A teacher writes on the board* or *A teacher teaches*. Ten items elicited 3rd sing. [-s] responses. For the past tense probe, participants were shown pictures of children engaged in activities, followed by a picture showing the activity being completed, and given prompts like *Here the boy is raking. Now he is done. Tell me what he did*. The expected answer would be *The boy/he raked*. Ten items elicited regular past tense [-ed] and eight items elicited irregular past tense forms.

In the BE/DO probe, children were prompted to ask questions or make statements using these morphemes. There were thirty-six items in total, twelve eliciting BE copula, thirteen BE auxiliary, and eleven DO. In this task, the child had to direct his or her questions to a puppet about one or more stuffed animals, or make statements about the animals. Thus, third singular and plural questions and statements were elicited using *is/are* and *do/does*. For instance, *I wonder if the bears are resting. Ask the puppet* was expected to prompt *Are the bears resting?* (BE auxiliary), and *I wonder about the Kitty. Ask the puppet if the Kitty is hungry* was expected to prompt *Is the Kitty hungry?* (BE copula).

In the TEGI GJ probe, the experimenter acted out a scenario with toys that includes two robots who children were told are just learning to speak English and do not say everything correctly. During the scenario, the children were asked to determine if the robots' statements were said correctly or incorrectly (*right* or *not so good*). There were thirty-five test items in this probe. The TEGI GJ probe examines children's ability to detect correct use or omission of verb morphemes, e.g. Dropped Marker: e.g. *he jumps over there/*he jump over there* or *he is jumping over there/*he jumping over there*, correct use or incorrect use of morphemes, e.g. Bad Agreement, e.g. *he is jumping/*he am jumping*, and omission of the progressive [-ing], e.g. Dropped ING, e.g. *he is jumping/*he is jump*.

For the production probes, a proportion correct score for the morpheme targeted was calculated by dividing the child's correct responses by the total of scorable responses. Unscorable responses are those that were imitating the experimenter's prompt, off-topic, or included a completely different verb construction, e.g. present progressive on the past tense probe. Because this study included older children with long-term exposure to English, unscorable responses were uncommon. For the past tense probe, the TEGI scoring procedures include accepting over-regularized past tense forms, e.g. *digged* for *dug*, as correct. For the purposes of this study, we scored over-regularizations as incorrect when we separated regular from irregular past in the analyses, but used the TEGI procedure

when comparing a child's score to the TEGI norms. For the GJ probe, following the Examiner's Manual, children's correct rejections, false alarms, misses, and hits were calculated and transformed into A-prime scores for Dropped Marker, Bad Agreement, and Dropped ING separately (Rice & Wexler, 2001; Rice, Wexler & Redmond 1999).

The Alberta Language Environment Questionnaire (ALEQ). The ALEQ (Paradis, 2011; <http://www.linguistics.ualberta.ca/CHESL_Centre/Questionnaires.aspx>) was administered to one parent, usually the mother, and with the aid of an interpreter if needed. This instrument includes questions on various topics, including current language use by family members in the home and the richness of the child's English environment. Information on language use in the home was gathered through questions such as *What language does the mother speak with the child?* or *What language does the child speak with the mother?* where answers were on 5-point rating scales from 0 [English never/mother tongue always] to 4 [English almost always/mother tongue almost never]. The proportion of overall English use in the home (input and output) was calculated from these rating scales for each family member. Richness of the English environment was determined by calculating the number of English-language activities the child engaged in, i.e. book-looking, TV/computer watching, reciting songs/rhymes, extracurricular activities, playing with English-speaking friends, and the frequency of these activities per week, to yield a proportional score from 0 to 1·0. Other information gathered from the ALEQ was used for inclusion criteria like AoA and length of exposure to English in school, and to gauge socioeconomic background (maternal education). English-use-in-the-home functioned as the variable measuring quantity of English input outside school, and English richness was the variable measuring quality of English input outside school. It is important to point out that quantity and quality cannot be entirely separated, and we did not have research questions comparing quantity versus quality of input, but instead considered both variables to be measures of environmental factors potentially influencing children's L2 abilities. Descriptives for the variables of English-use-at-home and English richness are given for each round in Table 1. Note that the mean use of English among family members ranged from .36 to .40 across the rounds, indicating that Chinese was being spoken more often than English among most families at home.

Comprehensive Test of Phonological Processing (CTOPP). The CTOPP non-word repetition subtest (Wagner, Torgesen & Rashotte, 1999) was administered to the children. This test includes a list of non-words that increase in length in syllables, played to the child from a CD, and the child was asked to repeat each one right after hearing it. The child's responses were recorded for later scoring. Following the CTOPP scoring procedures, children's non-word repetitions were scored as correct (each

sound repeated correctly) or incorrect (missing or substituted sounds). Raw scores were converted to standard scores, which are corrected for age, and the descriptives are in Table 1 for each round. For this subtest, the standard mean is 10, with 1 SD range of 7–13. Non-word repetition is a measure of verbal short-term memory (S. Gathercole, 2006), and constituted one of the child-internal predictor variables in this study.

Peabody Picture Vocabulary Test (PPVT-IV). The PPVT (Dunn & Dunn, 2007), a measure of receptive vocabulary size, was also administered to the children. Children were asked to point to an image out of an array of four images that best matched a word spoken by the experimenter. As with non-word repetition, raw scores were converted to standard scores, to correct for age, and descriptives are in Table 1. The PPVT standard mean is 100, with 1 SD range of 85–115.

Language input frequency. For the linear mixed regression analyses, we included language-level predictor variables, word frequency and allomorph type. The frequencies for the individual inflected verbs for the 3rd sing. [-s] and past tense probes were derived from the Edmonton ELL corpus and used in previous studies (Blom & Paradis, 2013; Blom *et al.*, 2012). The Edmonton ELL corpus consists of the spontaneous speech of native English-speaker research assistants and English L2 children (different from those in the present study) recorded and transcribed for other research purposes. The majority of words in the corpus comes from the research assistants. The Edmonton ELL corpus is relatively small (<500,000 words), but it is representative of the speech the children in this study hear because it is based on the oral speech of individuals in the Edmonton area. Moreover, word frequencies in this corpus have proven predictive of children's accuracy with L2 morphology in prior studies, while those from larger, less representative corpora did not (see Blom & Paradis, 2013, and Blom *et al.*, 2012, for more details). Frequencies were log-transformed and entered into the data frame for each item (verb) on the TEGI probe.

RESULTS

Change in scores over time

Figures 1 and 2 present a visual display of the change in scores across the three rounds for the TEGI production and grammaticality judgement probes, respectively. The mean scores with the standard deviations are presented in Table 2. We used linear mixed logistic regression modeling with the lme4 package (Bates, Maechler, Bolker & Walker, 2013) in the R statistical programming environment (R Core Team, 2013) in order to address research question (1) above concerning whether these children have reached a plateau in their morphological acquisition. Child, item, and round were random factors, with a random intercept for each item, and a

https://doi.org/10.1017/S0305000915000562 Published online by Cambridge University Press

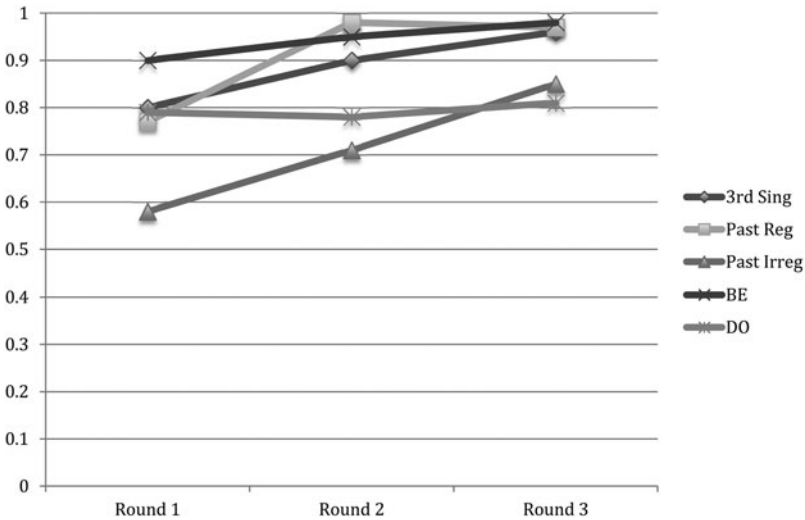

Fig. 1. Mean proportion correct scores for TEGI production probes across rounds.

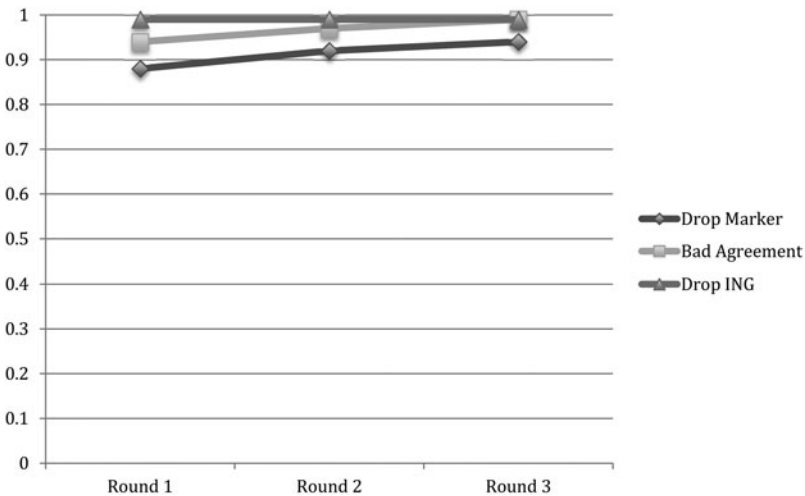

Fig. 2. Mean A-prime scores for TEGI grammaticality judgement probes across rounds.

random intercept and slope for each child dependent on round. Regarding random items, rather than aggregate proportion correct or A-prime scores for all items in a probe, for the modeling analysis, children's scores were coded as 'TRUE' or 'FALSE', indicating if the child gave a correct or

https://doi.org/10.1017/S0305000915000562 Published online by Cambridge University Press

TABLE 2. *Scores for TEGI probes across rounds and slope modeling results*

Probe and scores	Model 1: Round 1 to Round 2	Model 2: Round 2 to Round 3
3rd sing. [-s]	$z = 2{\cdot}56$, $p = {\cdot}0106$*	$z = 1{\cdot}85$, $p = {\cdot}0646$
R1 ·80(·30)		
R2 ·90(·24)		
R3 ·96(·07)		
Past regular	$z = 4{\cdot}60$, $p < {\cdot}0000$***	$z = -0{\cdot}03$, $p = {\cdot}7472$
R1 ·77(·38)		
R2 ·98(·05)		
R3 ·97(·06)		
Past irregular	$z = 2{\cdot}26$, $p = {\cdot}0240$*	$z = 1{\cdot}71$, $p = {\cdot}0868$
R1 ·58(·39)		
R2 ·71(·33)		
R3 ·85(·21)		
BE	$z = 3{\cdot}33$, $p = {\cdot}0009$***	$z = 2{\cdot}63$, $p = {\cdot}0085$**
R1 ·90(·12)		
R2 ·95(·08)		
R3 ·98(·03)		
DO	$z = 0{\cdot}167$, $p = {\cdot}8673$	$z = 0{\cdot}824$, $p = {\cdot}4101$
R1 ·79(·23)		
R2 ·78(·27)		
R3 ·81(·20)		
Dropped Marker	$z = 3{\cdot}06$, $p = {\cdot}0022$**	$z = 1{\cdot}24$, $p = {\cdot}216$
R1 ·88(·12)		
R2 ·92(·10)		
R3 ·94(·08)		
Bad Agreement	$z = 1{\cdot}97$, $p = {\cdot}0491$*	$z = 1{\cdot}81$, $p = {\cdot}0708$
R1 ·94(·09)		
R2 ·97(·07)		
R3 ·99(·03)		

NOTES: Scores are means (standard deviations), proportion correct or A-prime, for R1 (Round 1), R2 (Round 2), and R3 (Round 3). Individual items were entered into models, not proportion correct or A-prime scores. Model results are z-values followed by p-values for the fixed factor, Round. p-values < ·05 are marked with '*', p-values < ·01 with '**', and p-values < ·001 with '***'.

incorrect answer to each item within each TEGI probe. For these analyses, slopes between Round 1 to 2 were modeled (Model 1), and slopes between Round 2 to 3 were modeled (Model 2). If the fixed factor of Round was non-significant in both models, this would indicate a flat curve, i.e. no growth, across the three rounds. If Round was significant from Round 1 to 2 but non-significant from Round 2 to 3, this would indicate a plateau shape to the curve. Because accuracy was so stable and high for Dropped ING, no model was generated for this probe. The z and p values from these analyses are presented in Table 2 for each probe. Results indicate that for the DO probe, no change was found over time in the children's scores, i.e. flat curve. For the 3rd sing. [-s], past regular, past irregular, GJ Dropped Marker, and GJ Bad Agreement probes, change was found from

https://doi.org/10.1017/S0305000915000562 Published online by Cambridge University Press

Round 1 to 2, but not from Round 2 to Round 3, indicating a plateau shape in the curve. For the BE probe, continued change was found because significant change between rounds was found in both models.

Comparisons with monolingual criterion scores

TEGI is a criterion-referenced test. This means that individual scores are compared to a criterion or cut-off score, rather than converted to standard scores or percentiles, to assess whether an individual child's performance meets age expectations. TEGI criterion scores are the lowest possible score obtained by children with typical language development in the norming sample. (The TEGI norming sample included children with specific language impairment.) Each child's score was age-referenced to the appropriate criterion score from the TEGI Examiner's Manual; however, it is important to note that the monolingual typically developing children in the TEGI norming sample showed stable, ceiling performance from about 6;0 until 8;11 on all probes (Rice & Wexler, 2001). At the final round, the children in this study were older than the oldest group from the TEGI norming sample (8;11), and so we used the last criterion score given. This was not problematic because the monolingual children's ceiling scores had not changed over three years and there is no plausible reason to believe their scores would have gone down after that point. One final explanation regards the past tense probe. As mentioned in 'Method', the TEGI scoring procedures assign a correct score for a past irregular form that is over-regularized, e.g. *digged* instead of *dug* would be correct because it has morphological marking. This is why past regular and irregular are not separated for this analysis.

Because we were interested in children's long-term outcomes (see research question (2) above), we focused on Round 3 for this comparison with monolinguals. We assigned children a score of 1 (score is at or above criterion score) or 0 (score is below criterion) for each TEGI production and grammaticality judgement probe. Table 3 shows the individual criterion scores for each of the children at Round 3, along with the means and SDs for each probe. Note that for the probe means, 1·0 indicates all children reached criterion; a mean of 0 indicates no child reached criterion. Table 3 shows that 11/18 or 61% of the children had scores below criterion on one or more probes. Note that both Mandarin L1 and Cantonese L1 children are represented among children not meeting criterion scores on the TEGI. Of the 11 children not meeting TEGI criterion scores, none had met the criterion at a previous round for that probe (i.e. no backsliding). Of the 11 children with scores below criterion, 4 had 1 score below, 3 had 2 scores below, 3 had 3 scores below and 1 had 4 scores below. Regarding the probes, the only probe where all children

TABLE 3. *Individual TEGI criterion scores for probes at Round 3*

Child	3rd sing.	Past	BE	DO	Dropped marker	Bad agreement	Dropped ING
01MA	1	1	1	0	0	1	1
02MA	1	1	1	1	1	1	1
03MA	1	1	1	0	1	1	1
04MA	1	1	1	0	1	1	1
05MA	1	1	1	1	1	1	1
06CA	0	1	1	0	0	1	1
07CA	1	0	1	0	0	1	1
08CA	0	0	1	1	1	1	1
09CA	1	1	1	1	1	1	1
10MA	1	1	1	0	1	1	1
11CA	0	1	0	0	0	1	1
12CA	1	1	1	0	0	1	1
13CA	1	1	1	1	1	1	1
14MA	1	1	1	1	1	1	1
15CA	1	1	1	1	1	1	1
16CA	1	0	1	0	1	0	1
17CA	0	1	1	1	1	1	1
18MA	1	1	1	1	1	1	1
Mean	0·78	0·83	0·89	0·50	0·72	0·94	1·00
SD	0·43	0·38	0·32	0·51	0·46	0·24	0·00

NOTES: 'MA' = Mandarin L1; 'CA' = Cantonese L1. '1' means the child's score for that probe was at or above the criterion score for their age, based on monolingual norms. '0' means the child's score was below the criterion score.

reached criterion was Dropped ING. For BE and Bad Agreement, just 1 child did not reach criterion on these probes, but it was a different child for each probe. For the past tense, 3 children did not meet criterion, for 3rd sing. [-s], 4 children, Dropped marker, 5 children, and DO, 9 children.

All the children in this study began to learn English before the age of 6;0, but there was still variation in their AoAs. We next examined whether children who did not meet criterion scores for one or more TEGI probes had older AoAs than the children who did meet the criterion scores. The 11 children who did not meet criterion on one or more TEGI probes began to learn English between ages 1;7 and 5;5 (M = 4;2). The 7 children who did meet the criterion scores on all the TEGI probes began to learn English between ages 3;8–5;8 (M = 4;2). Thus, younger age of English onset did not seem associated with whether children reached native-speaker levels of performance on the TEGI.

The children had a mean exposure of 6;4 to English at Round 3, but, as with AoA, there was some individual variation in length of exposure at each round. We next examined whether children who did not meet criterion scores on the TEGI at Round 3 had less exposure to English than the children who did meet the criterion scores. The 11 children who did not meet criterion had between 5;5–7;4 years of exposure to English

(M = 6;5); the 7 children who did meet the criterion had between 5;7–7;1 years of exposure (M = 6;2). Therefore, also as with AoA, differences in exposure time to English among the children at Round 3 did not appear to explain whether or not children's scores met the criterion on the TEGI.

Individual difference and language-level factors

We used linear mixed logistic regression modeling in R to determine which factors most influenced children's performance on each probe across the three rounds. Child, item, and round were random factors, with a random intercept for each item, and a random intercept and slope for each child dependent on round. Fixed factors included child-level (individual difference) and item-level (language-level) factors. Child-level factors were: English-use-at-home (input quantity), English Richness (input quality), CTOPP (non-word repetition – verbal short-term memory), and PPVT (vocabulary size). Item-level factors were word frequency (3^{rd} sing. [-s], past regular and irregular only) and allomorph (3^{rd} sing. [-s] and past regular only).

For the child-level factors, correlations were performed between the values at each round to determine if any were .75 or higher. The highest correlation was .53, observed at Round 3 between CTOPP and PPVT, thus none were high enough for co-linearity to be an issue. The moderate correlation between these factors is not surprising given the well-established relationship between vocabulary size and verbal short-term memory in children (S. Gathercole, 2006). Both factors were entered into the models because, in spite of this relationship, they measure separate constructs, and thus could exert separate influences on children's performance with verb morphology. Moreover, if a correlation in a sample also exists for the entire population, this correlation is not expected to be problematic for a model (Harrell, 2001).

Because there were just 18 children in our sample, the final optimal model for each probe was restricted to two child-level fixed factors, following the convention of approximately one factor to ten participants. There were just two item-level factors, allomorph and word frequency. Because each item was considered individually, the number of items was sufficient such that no restrictions on the number of item-level factors for the probes of verb inflection were required. Because we considered four child-level factors in total in our study (English-use-at-home, English richness, CTOPP, and PPVT), we had a three-step process to determine the optimal model for each TEGI probe. The first step was to generate the best full model for each probe, i.e. the best-fitting model (lowest deviance as measured by AIC) with a maximum of two child-level factors, and any number of the item-level factors. So, the child-level factors were entered

systematically, in pairs, to generate several models in order to find the best-fitting one. The second step was to determine if the best-fitting full model for each probe was superior to a reduced model with one child-level or item-level factor removed at a time (the difference in deviance produced a significant chi-square value at 1 degree of freedom: $X^2 > 3{\cdot}84$, $p < {\cdot}05$). If so, the full model was chosen; if not, the reduced model was chosen as the optimal model. Step three consisted of calculating the Concordance Index *C* for the optimal model to assess whether this optimal model was a good fit. *C* ranges from .50 to 1·0, and models of .80 or higher are considered to be good-fitting models (Chatterji & Hadi, 2006). As with the analyses of curve shape over time, because accuracy was so stable and high for Dropped ING, no model was produced for this probe.

The summary results of this process are in Table 4. For each TEGI probe, the significant fixed factors and *C* for the optimal model are listed. For the DO probe, two models are given because they produced nearly equal deviances, and thus a best-fitting model could not be determined between them. For the 3rd sing. [-s] probe, the optimal model of children's performance included PPVT ($z = 3{\cdot}25$, $p = {\cdot}0012$) and English richness ($z = 3{\cdot}051$, $p = {\cdot}0023$) as child-level predictors, indicating that children with bigger English vocabularies and richer English environments outside school had greater accuracy with this morpheme in production. One item-level predictor, allomorph (allomorph-s: $z = 2{\cdot}17$, $p = {\cdot}0302$; allomorph-z: $z = 2{\cdot}89$, $p = {\cdot}0039$) was significant, indicating that children were more accurate in adding the 3rd sing. [-s] with verbs taking the allomorphs [-s] and [-z] than the allomorph [-ɪz]. For past regular, children's performance was best predicted by CTOPP ($z = 2{\cdot}92$, $p = {\cdot}0036$) and PPVT ($z = 2{\cdot}25$, $p = {\cdot}0247$) and allomorph (allomorph-ɪd: $-2{\cdot}84$, $z = -2{\cdot}52$, $p = {\cdot}0119$). Thus, children with bigger vocabularies and superior verbal short-term memories were more accurate with regular past tense marking. For the allomorph variable, children were less accurate in producing the past tense with verbs taking the allomorph [-ɪd] than with verbs taking [-d] or [-t]. The optimal model for the past irregular forms included both CTOPP ($z = 3{\cdot}40$, $p < {\cdot}0000$) and PPVT ($z = 2{\cdot}65$, $p = {\cdot}0079$) as predictors. As with past regular, children with bigger vocabularies and superior verbal short-term memories were more accurate with past irregular. In addition, the frequency of the target verb appearing in its irregular past tense form in the input predicted more accurate use of this form by the children ($z = 2{\cdot}56$, $p = {\cdot}0105$). For the BE probe, children's responses were best predicted by CTOPP/verbal short-term memory ($z = 3{\cdot}77$, $p = {\cdot}0002$) and PPVT/vocabulary size ($z = 5{\cdot}97$, $p < {\cdot}0000$). Regarding the DO probe, the first model included CTOPP ($z = 2{\cdot}61$, $p = {\cdot}0092$) and English-use-at-home ($z = 3{\cdot}54$, $p = {\cdot}0004$) as predictors, and the second model included PPVT ($z = 2{\cdot}45$, $p = {\cdot}0141$) and

TABLE 4. *Optimal logistic regression models for each TEGI probe*

TEGI probe	Significant fixed factors in optimal model	Concordance index
3rd sing. -s	PPVT + English richness + Allomorph (-s, -z, -ɪz)	0·94
Past-regular	CTOPP + PPVT + Allomorph (-t, -d, ɪd)	0·98
Past-irregular	CTOPP + PPVT + Word frequency	0·94
BE	CTOPP + PPVT	0·95
DO	CTOPP + English-use-at-home	0·90
	PPVT + English-use-at-home	0·91
Dropped Marker	CTOPP + PPVT	0·97
Bad Agreement	PPVT	0·93

NOTES: Child-level factors are PPVT (vocabulary size), English richness (richness of the English environment outside school), CTOPP (verbal short-term memory), and English-use-at-home (proportional use of English among family members). Item-level factors are allomorph (allomorphs required by verbs on TEGI probes), and word frequency (frequency of the inflected word form in the input).

English-use-at-home ($z = 2{\cdot}41$, $p = {\cdot}0160$). Thus, children were more accurate with DO forms when they heard/spoke more English at home and if they had bigger vocabularies or superior verbal short-term memories. Turning to the grammaticality judgement probes, for the Dropped Marker probe, children's responses were best predicted by a model including CTOPP/verbal short-term memory ($z = 2{\cdot}84$, $p = {\cdot}0045$) and PPVT/vocabulary ($z = 2{\cdot}76$, $p = {\cdot}0058$). For Bad Agreement, the most parsimonious optimal model included just PPVT/vocabulary ($z = 4{\cdot}70$, $p < {\cdot}0000$).

DISCUSSION

This longitudinal study examined the long-term outcomes with English L2 verb morphology of Chinese L1 children who all began to learn English in early childhood ($\text{AoA}_{\text{mean}} = 4;2$). Our research questions were aimed at determining (1) if children's developmental trajectories were slowing down/reaching a plateau, (2) if children had reached native-speaker levels of accuracy with the morphemes, and (3) what individual-difference and language-level factors played a role in shaping their abilities with L2 morphology during this late stage of their acquisition.

Developmental trajectories and native-like attainment

Two hallmark characteristics of English monolingual morphological acquisition are that growth reaches a plateau over time (by age 6;0) and that accuracy reaches ceiling at asymptote, with very little individual variation at that point (Rice & Wexler, 1996; Rice, Wexler & Hershberger, 1998). These characteristics are evident in the TEGI norming sample

(Rice & Wexler, 2001). Our analyses examined growth over time with L2 morphology and an interpretation of L2 children's outcomes in terms of their monolingual peers.

Overall, the L2 children's proportion correct and A-prime scores on the TEGI were highly accurate by Round 3, where mean scores for all probes were >·80, and for all probes except past irregular and DO, they were >·90. The analyses of developmental trajectories indicated that for the majority of probes, the children were showing a plateau in their growth in performance by the end of the study, since no changes emerged between slopes for Rounds 2 and 3 except for the BE probe. These results point to the possibility that children's development of L2 morphology could be approaching asymptote or, at least, slowing down. In addition, not all the children reached native-speaker levels of accuracy on all the probes, but instead there was variation in outcomes across children and probes. Regarding individual differences among the children, 61% did not meet native-speaker levels of accuracy with one or more probes after 6;4 years of exposure, but importantly, 39% of the children did converge on native-speaker performance for all the probes. Thus, the likelihood of attaining native-speaker accuracy within the time period of this study varied among individuals in spite of all of them having early AoAs. Differences among the probes emerged with respect to children's convergence on native-speaker abilities, ranging from the Dropped -ING probe where 100% of the children reached native-speaker accuracy levels to the DO probe where just 50% of the children met native-speaker levels of accuracy. These results point to how children seemed to be acquiring these morphological constructions not as a group but rather on a piecemeal basis. Furthermore, a few children did not reach criterion on the Dropped Marker GJ probe, indicating that divergence with native-speaker abilities was not merely a production problem.

A pattern emerged across these results suggesting that verb inflection, particularly inflection marking subject–verb agreement (3^{rd} sing. -s, DO, Dropped Marker), could be exceptionally problematic for Chinese L1 learners of English. First, the superior long-term attainment with the BE probe in production versus the production probes involving inflectional verb morphology, past tense and 3^{rd} sing. [-s], could be expected given that English L2 children's accuracy with BE outpaces their accuracy with verb inflections at earlier stages of acquisition, and Chinese L1 children are slower to acquire verb inflection in L2 English than children from other L1 backgrounds (Blom *et al.*, 2012; Paradis, 2008; Paradis, Rice, Crago & Marquis, 2008). Second, the most common error children made on the DO probe was use of the plural 'DO' when the third singular 'DOES' was required, rather than omission of DO, also signalling a problem with agreement inflection. Furthermore, half of the stimuli on the

Dropped Marker probe involved omission of 3rd sing. [-s]. An examination of children's scores for stimuli targets with dropped BE vs. verb inflection on this probe indicate higher scores for the former at Round 3 ($t = 2{\cdot}71$, $p = {\cdot}016$). Finally, the three probes involving verb inflection related to agreement, 3rd sing. [-s], DO, and Dropped Marker, were the probes for which the largest number of children did not meet criterion by Round 3. That morphology related to subject–verb agreement would be particularly problematic is consistent with Tsimpli's (2014) proposal that semantically vacuous, narrow syntax is more vulnerable to delayed AoA than other grammatical subdomains.

How do these results compare to other studies of long-term outcomes with English L2 verb morphology? Jia and Fuse (2007) found that developmental trajectories of correct use of morphology in spontaneous speech accelerated at first, but reached plateau by the end of five years, paralleling our results. Jia and Fuse found a great deal of variation among children and morphemes, also in parallel with the present study. Regarding native-like attainment, Jia and Fuse used 80% correct use in context as a criterion for 'mastery', and found that the early AoA children were more likely to reach mastery for morphology than the older AoA children. If the more conventional criterion of 90% correct use in spontaneous speech (e.g. Brown, 1973) were applied to their data, it would appear that some of their early AoA children did not reach this more stringent criterion for mastery with 3rd sing. [s], DO, BE, past regular and irregular by the end of five years of exposure. Marinis and Chrondrogianni (2010) compared Turkish-L1–English-L2 children's performance on standardized tests of English, including the TEGI, to the performance of monolingual age peers. While this was not a study focused on the influence of AoA on child L2 acquisition, they do report some relevant long-term attainment findings on the TEGI, as the L2 children had an average of four years of exposure to English and AoAs < 6;0. Marinis and Chondrogianni found no differences between the L2 and monolingual children on the TEGI past tense probe, but the L2 children performed worse than monolinguals on the 3rd sing. [-s] probe. Their analysis of individual scores indicated that 5/6 nine-year-old children with six years of exposure did meet age-expected criterion scores for this probe. Thus, Marinis and Chrondrogianni also found variation between probes and individuals, but it seems that convergence with native-speaker performance could be expected for Turkish L1 children by six years of exposure.

This difference between Marinis and Chondrogianni (2010) on one hand, and Jia and Fuse (2007) and the present study on the other, might be due to Turkish being an inflected language and Cantonese and Mandarin being isolating languages. Speakers of isolating L1s like Cantonese, Mandarin, and Vietnamese show greater difficulty in acquiring verb inflection than

speakers of languages with rich inflection, like Spanish or Punjabi, in both the early stages and in long-term outcomes (Blom *et al.*, 2012; McDonald, 2000; Paradis, 2011). The particular vulnerability of inflectional morphology in English as an L2 could be a combination of the following: (i) low saliency and cue reliability of morphology in the input; (ii) the filter of L1 phonological constraints on codas and consonant clusters; and (iii) the need to re-focus attention and processing routines in order to acquire grammatical features not present in the L1 (Blom *et al.*, 2012; Ellis, 2008; Flege *et al.*, 1999; Sorenson Duncan & Paradis, in press). Even though low saliency and cue reliability would affect all English L2 learners, the phonological and morphological characteristics of Mandarin and Cantonese would render verb inflection more challenging for speakers of these L1s because they cannot benefit from positive L1 transfer. The variability in long-term outcomes with morphology found in the present study points to the possibility of long-lasting effects of L1 influence even in L2 learners with early AoAs. However, because all the children had a Chinese L1 in the present study, this conjecture needs to be tested with further research including children from other L1 backgrounds.

The present study was designed to look at long-term child L2 outcomes, and it is relevant to ask whether these long-term outcomes might signal children's ultimate L2 attainment. Recall that developmental retrospective studies with adults have found that early AoA L2 speakers do not always converge on native-speaker grammatical abilities, including verb morphology (Abrahamsson & Hyltenstam, 2009; Flege *et al.*, 1999; McDonald, 2000; Weber-Fox & Neville, 1999). Thus, divergence is a possible long-term outcome for the L2 children in this study. Recall also that monolingual children reach ceiling on the TEGI probes by age 6;0, and some child L2 speakers in this study did not converge on native-speaker accuracy even after six years of exposure to English. Furthermore, the shape of the developmental trajectories suggests that the L2 children might not get much closer to native-speaker accuracy in the future, for the DO probe in particular. However, for most probes, a plateau was only evident between Rounds 2 and 3, and a finding of non-native levels of accuracy extending over a longer period of time would constitute more convincing evidence for these speakers having reached their ultimate attainment in the L2. At the limit, we believe this study suggests that even early AoA child L2 learners could be AT RISK for divergence from monolinguals in their accuracy with English morphology, and that this divergence might be evident in their L2 by four to six years of exposure. Further research with childhood L2 learners with even longer exposure would be needed to draw conclusions about ultimate attainment with certainty.

Non-age predictors of L2 morphological acquisition

The variation observed with respect to children's individual outcomes on the TEGI probes indicate that other, non-AoA, factors were influencing their acquisition, and the results of our linear mixed regression analyses confirm this indication. We found that the child-internal factors, verbal short-term memory (CTOPP), and vocabulary size (PVVT) were the most common predictors, and the environmental factors, English richness, and English-use-at-home appeared less frequently in the models. However, it is notable that English richness did emerge as a significant predictor in other good-fitting but not optimal models for past regular, past irregular, and BE. English-use-at-home was also a significant predictor in a good-fitting model for the BE responses. Thus, while the internal factors were exerting the strongest influence, the influence of environmental factors on children's performance was also present. Regarding language-level factors, we found that allomorph type influenced accuracy with 3rd sing. [-s] and past regular, and word frequency influenced accuracy with past irregular. Because factors like superior verbal short-term memory or richer English input predict higher scores on the TEGI, in turn they predict whether children achieve native-like abilities or not because higher scores are more likely to reach the age-expected criterion. As such, the results of this analysis show that non-AoA factors could play a decisive role in whether or not early L2 learners catch up to their native-speaker peers with L2 verb morphology after four to six years of exposure.

It is relevant to consider whether the secondary role of input factors could have been an artifact of how they were measured in the present study, because English-use-at-home and English richness are composite and indirect (parent report) measures. Prior research indicates that input factors such as diversity of speakers, family composition, and parents' fluency in the L2 can exert independent influences on bilingual children's development (Armon-Lotem *et al*., 2014; Hoff *et al*., 2014). In this study, English-use-in-the-home is a composite measure of both input to children and children's output. Is it possible that parents' use of English to the children, if that English were heavily accented and contained morphological errors, could have contributed to the variability in children's English output? The proportion of English-use-in-the-home among all family members was .36 to.40 on average (see Table 1), meaning more Chinese was spoken amongst them. Breaking apart the variable of English-use-in-the-home to just parents' use of English, the proportion shrinks to .27–.28 on average across three rounds. It is also important to keep in mind that the city itself, and the schools children were attending, are culturally and linguistically diverse, and so children's input outside the home was comprised of a variety of English speakers,

https://doi.org/10.1017/S0305000915000562 Published online by Cambridge University Press

including native speakers. Furthermore, it is reasonable to assume that for eight- to ten-year-old children, the variety of language input sources (school, friends, media) beyond interaction with parents would be much greater than for younger children. Therefore, the small amount of individual variation predicted by English-use-at-home, together with the limited amount of English actually used by parents at home, indicate that for this sample of children, it is unlikely that parents' accented speech was a major contributor to the variability in children's outcomes. Nevertheless, future research examining input factors and children's L2 outcomes should include a more fine-grained breakdown of these factors and some direct measures. Doing so would enable us to better understand the balance between internal and external factors predicting individual variation in morphological acquisition.

Finally, the analyses in this study revealed that factors influencing children's L2 abilities with verb morphology at the early stages of acquisition continue to shape their development even at later stages. For example, studies with child L2 learners with less L2 exposure have found that verbal short-term memory, vocabulary size, L2 input quality and quantity, and allomorph and word frequency influence L2 morphological acquisition (Armon-Lotem *et al.*, 2011; Blom & Paradis, 2013, 2015; Blom *et al.*, 2012; Paradis, 2011). Other studies that included children with long-term exposure have also found that L2 input quality and quantity factors (Chrondrogianni & Marinis, 2011; Jia & Fuse, 2007, Unsworth, 2013; Unsworth *et al.*, 2014), and language-level factors (Marinis & Chrondrogianni, 2010) predict L2 grammatical abilities. The results from this study, together with the existing literature, raise the question of why individual difference and language-level factors would still be exerting an effect at later stages in acquisition. For example, if a vocabulary of a certain size constituted a 'critical mass' needed for children to begin to become productive with verb morphology (e.g. Marchman & Bates, 1994), why would vocabulary size matter at later stages when productivity is clearly evident? The continued influence of these individual difference and language factors across years of L2 acquisition indicate that they do not serve a kind of 'bootstrapping' function early on and then fade away. The continued influence of these factors could be argued to support Usage-Based or Emergentist models of morphology and the lexicon (e.g. Bybee, 2010; Ellis, 2008). This is because such models assume that lexical composition, input frequency, and cognitive mechanisms like verbal memory skills all impact morphological learning, processing, and use across the lifespan, and also influence diachronic change.

CONCLUSIONS

This study found that there was individual variation in the children's long-term L2 outcomes with verb morphology. Thirty-nine percent of the children had acquired native-like levels of accuracy for all morphemes by 6;4 years of L2 exposure, whereas 61% had not reached this level for all morphemes. Because all the children in this study had similar early AoAs, the likelihood of individuals achieving native-like accuracy in their L2 in this timeframe was due to non-AoA factors. Variation in children's accuracy with English morphology was predicted by variation in verbal short-term memory, vocabulary size, and child- and language-level input factors. Our results suggest that the four to six years 'catching up' timeframe for L2 oral language (Hakuta *et al*., 2000; Saunders & O'Brien, 2006) is insufficient for verb morphology, at least for children from typologically isolating L1 backgrounds. Another way to interpret these results is that this timeframe is sufficient because variable use, and thus divergence with monolingual accuracy levels, might constitute the long-term outcomes for some bilingual speakers. This alternative interpretation raises the broader question of what the appropriate expectations are for child L2 acquisition, and whether monolingual speakers should be the 'gold standard' for comparison (Ellis, 2008; Muñoz & Singleton, 2011). Child bilingual speakers arguably have more sources of variation in their learning experience than monolinguals, and since these sources of variation shape their L2 acquisition at both early and later stages, it is logical to expect greater variability in linguistic outcomes. While we believe comparisons between monolingual native speakers and bilinguals can be informative from a scientific perspective, we also believe that interpretations of divergence between child bilinguals and monolinguals in long-term outcomes should be careful not to promote a deficit view of bilingualism (cf. Muñoz & Singleton, 2011).

REFERENCES

Abrahamsson, N. & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: listener perception versus linguistic scrutiny. *Language Learning* **59**, 249–306.

Armon-Lotem, S., Joffe, S., Abutbul-Oz, H., Altman, C. & Walters, J. (2014). Language exposure, ethnolinguistic identity and attitudes in the acquisition of Hebrew as a second language among bilingual preschool children from Russian- and English-speaking backgrounds. In T. Grüter and J. Paradis (eds), *Input and experience in bilingual development*, 77–98. Amsterdam: Benjamins.

Armon-Lotem, S., Walters, J. & Gagarina, N. (2011). The impact of internal and external factors on linguistic performance in the home language and in L2 among Russian–Hebrew and Russian–German preschool children. *Linguistic Approaches to Bilingualism* **1**, 291–317.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2013). lme4: linear mixed-effects models using Eigen and S4. R package version 1.0–5. Online: <http://CRAN.R-project.org/package=lme4>.

Blom, E. & Paradis, J. (2013). Past tense production by English second language learners with and without impairment. *Journal of Speech, Language and Hearing Research* **56**, 1–14.
Blom, E. & Paradis, J. (2015). Sources of individual differences in the acquisition of tense inflection by English second language learners with and without specific language impairment. *Applied Psycholinguistics* **36**, 953–76.
Blom, E., Paradis, J. & Sorenson Duncan, T. (2012). Effects of input properties, vocabulary size and L1 on the development of third person singular -s in child L2 English. *Language Learning* **62**, 965–94.
Bohman, T., Bedore, L., Peña, E., Mendez-Perez, A. & Gillam, R. (2010). What you hear and what you say: language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism* **13**, 325–44.
Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press.
Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
Chatterjee, C. & Hadi, A. S. (2006). *Regression analysis by example*. New York: Wiley.
Chondrogianni, V. & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism* **1**, 318–42.
Conboy, B. & Thal, D. (2006). Ties between the lexicon and grammar: cross-sectional and longitudinal studies of bilingual toddlers. *Child Development* **77**, 712–35.
DeKeyser, R. (2012). Age effects in second language learning. In S. Gass & A. Mackey (eds), *The Routledge handbook of second language acquisition*, 442–60. New York: Routledge.
Dörnyei, Z. & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (eds), *The handbook of second language acquisition*, 589–630. Oxford: Blackwell.
Dunn, L. & Dunn, D. (2007). *Peabody Picture Vocabulary Test–4th edition*. San Antonio, TX: Pearson.
Ellis, N. (2008). The dynamics of second language emergence: cycles of language use, language change and language acquisition. *Modern Language Journal* **92**, 232–49.
Flege, J., Munro, M. & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a language. *Journal of the Acoustical Society of America* **97**, 3125–38.
Flege, J., Yeni-Komshian, G. & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory and Language* **41**, 78–104.
Gathercole, S. E. (2006). Keynote article: nonword repetition and word learning: the nature of the relationship. *Applied Psycholinguistics* **27**, 513–43.
Gathercole, V. M. (2007). Miami and North Wales, so far and yet so near: a constructivist account of morpho-syntactic development in bilingual children. *International Journal of Bilingual Education and Bilingualism* **10**, 224–47.
Goldschneider, J. & DeKeyser, R. (2001). Explaining the 'natural order of L2 morpheme acquisition' in English: a meta-analysis of multiple determinants. *Language Learning* **51**, 1–50.
Hakuta, K., Bialystok, E. & Wiley, E. (2003). Critical evidence: a test of the critical period hypothesis for second language acquisition. *Psychological Science* **14**, 31–8.
Hakuta, K., Goto Butler, Y. & Witt, D. (2000). *How long does it take English learners to attain proficiency?* Policy Report, the University of California Linguistic Minority Research Institute. Online: <http://www.stanford.edu/~hakuta/>.
Harley, B. & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition* **19**, 379–400.
Harrell, F. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer-Verlag.
Hoff, E., Welsh, S., Place, S. & Ribot, K. (2014). Properties of dual language input that shape bilingual development and properties of environments that shape dual language input. In T. Grüter & J. Paradis (eds), *Input and experience in bilingual development*, 119–40. Amsterdam: Benjamins.

Jia, G. & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics* **24**, 131–61.
Jia, G. & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents. *Journal of Speech, Language and Hearing Research* **50**, 1280–99.
Lin, H. (2001). *A grammar of Mandarin Chinese*. Muenchen: Lincom Europa.
Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.
Marchman, V. A. & Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language* **12**, 339–66.
Marchman, V. A., Martínez-Sussmann, C. & Dale, P. S. (2004). The language-specific nature of grammatical development: evidence from bilingual language learners. *Developmental Science* **7**, 212–24.
Marinis, T. & Chondrogianni, V. (2010). Production of tense marking in successive bilingual children: When do they converge with their monolingual peers? *International Journal of Speech-Language Pathology* **12**, 19–28.
Marinova-Todd, S. H., Marshall, D. B. & Snow, C. E. (2000). Three misconceptions about age and second language acquisition. *TESOL Quarterly* **34**, 9–34.
Masoura, E. V. & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology* **34**, 383–8.
Matthews, S. & Yip, V. (1994). *Cantonese: a comprehensive grammar*. New York: Routledge.
McDonald, J. (2000). Grammaticality judgments in a second language: influences of age of acquisition and native language. *Applied Psycholinguistics* **21**, 395–423.
McDonald, J. (2008). Grammaticality judgments in children: the role of age, working memory and phonological ability. *Journal of Child Language* **35**, 247–68.
Meisel, J. (2008). Child second language acquisition or successive first language acquisition? In E. Gavruseva & B. Haznedar (eds), *Current trends in child second language acquisition: a generative perspective*, 55–80. Amsterdam: Benjamins.
Meisel, J. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft* **28**, 5–34.
Montrul, S. (2008). *Incomplete acquisition in bilingualism: reexamining the age factor*. Amsterdam: John Benjamins.
Muñoz, C. & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching* **44**, 1–35.
Paradis, J. (2008). Tense as a clinical marker in English L2 acquisition with language delay/impairment. In E. Gavruseva & B. Haznedar (eds), *Current trends in child second language acquisition: a generative perspective*, 337–56. Amsterdam: Benjamins.
Paradis, J. (2011). Individual differences in child English second language acquisition: comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism* **1**, 213–37.
Paradis, J., Rice, M., Crago, M. & Marquis, J. (2008). The acquisition of tense in English: distinguishing child L2 from L1 and SLI. *Applied Psycholinguistics* **29**, 1–34.
R Core Team (2013). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Online: <http://www.R-project.org/>.
Rice, M. L. & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment. *Journal of Speech, Language, and Hearing Research* **39**, 1236–57.
Rice, M. L. & Wexler, K. (2001). *Test of Early Grammatical Impairment*. New York: The Psychological Corporation.
Rice, M. L., Wexler, K. & Hershberger, S. (1998). Tense over time: the longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* **41**, 1412–31.
Rice, M. L., Wexler, K. & Redmond, S. (1999). Grammaticality judgments of an extended optional infinitive grammar: evidence from English-speaking children with SLI. *Journal of Speech, Language, and Hearing Research*, **42** 943–61.

Saunders, W. & O'Brien, G. (2006). Oral language. In F. Genesee, K. Lindholm-Leary, W. Saunders & D. Christian (eds), *Educating English language learners: a synthesis of research evidence*, 14–63. Cambridge: Cambridge University Press.

Simon-Cereijido, G. & Gutiérrez-Clellen, V. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexicon and grammar. *Applied Psycholinguistics* **30**, 315–37.

Sorenson Duncan, T. & Paradis, J. (in press). English language learners' nonword repetition performance: the influence of L2 vocabulary size, length of L2 exposure and L1 phonology. *Journal of Speech, Language, and Hearing Research*.

Statistics Canada (2011). *Linguistic characteristics of Canadians*. Catalogue no. 98-314-X2011001. Online: <http://www12.statcan.ca/census-recensement/2011/as-sa/98-314-x/98-314-x2011001-eng.cfm>.

Tsimpli, I. M. (2014). Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism* **4**, 283–313.

Unsworth, S. (2013). Assessing age of onset effects in (early) child L2 acquisition. *Language Acquisition* **20**, 74–92.

Unsworth, S., Argyri, F., Cornips, L., Hulk, A. C. J., Sorace, A. & Tsimpli, I. (2014). On the role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics* **35**, 765–805.

Wagner, R., Torgesen, J. & Rashotte, C. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

Weber-Fox, C. & Neville, H. (1999). Functional neural subsystems are differentially affected by delays in second language immersion: ERP and behavioral evidence in bilinguals. In D. Birdsong (ed.), *Second language acquisition and the critical period hypothesis*, 23–38. Mahwah, NJ: Erlbaum.