The development of English tense and agreement morphology in Welsh-English bilingual children with and without specific language impairment

The development of English tense and agreement morphology in Welsh-English bilingual children with and without Specific Language Impairment

Running title: Acquisition of English tense in Welsh children

Abstract

This study investigated whether third person singular (3SG) *–s* and past tense accuracy and error types can reveal distinct developmental patterns of agreement and tense acquisition in younger and older Welsh L1 – English sequential bilingual (L2) children with typical development (L2-TLD) and in younger children with language impairment (L2-SLI_Y). A group of older (L2-TLD_O) (mean age: 93.72 months) and younger (L2-TLD_Y) (mean age: 67 months) Welsh-English typically developing bilingual children and a group of young (mean age: 63 months) children with SLI (L2-SLI_Y) age-matched to the L2-TLD_Y group were administered the screening component of the Test of Early Grammatical Impairment (Rice & Wexler, 2001). The results indicated that the three groups differed in their production of 3SG *–s* and regular past tense but not in terms of accuracy on irregular past tense verbs, when vocabulary skills were considered. The L2-SLI_Y children produced similar error types to the L2-TLD_Y children, who differed from their L2-TLD_O peers in this respect. L2 children's vocabulary size, non-verbal intelligence and item-level factors, such as frequency and morphophonology, differentially contributed to their performance across the various morphemes. We discuss these results within current accounts of language development and impairment.

Key Words: child second language acquisition, specific language impairment, third person singular, regular and irregular past tense.

Introduction

Experimental studies with sequential bilingual (L2) children and especially those investigating inflectional morphology usually focus on how school-age L2 children with typical language development (TLD) (between the ages of 7 and 9 years) in mainstream English education and in countries where English is the language of the community acquire morphological markers. Although these studies offer a snapshot of children's morphological acquisition at that specific age (Chondrogianni & Marinis, 2011; Marinis & Chondrogianni, 2010; Paradis, 2011) and highlight how child-internal factors (age of acquisition, length of exposure, proficiency) and language- or item-specific properties (morphophonology, frequency) influence the acquisition outcomes (Chondrogianni & Marinis, 2011; Marinis & Chondrogianni, 2010; Paradis, 2011), they do not inform us about how accuracy and error types in L2 children change over time as a function of child-internal and language-specific properties. To date, only two studies by Paradis and colleagues have investigated how tense and agreement develop over time from the age of 5 to 7 (Blom, Paradis & Duncan, 2012) and from the age of 8 to 10 years (Paradis, Jia & Arppe, 2016) in children from diverse first language (L1) backgrounds (Blom et al., 2012) or in children with the same L1 (Chinese) (Paradis et al., 2016). The present study builds on this line of research by examining how Welsh L1 – English L2-TLD children attending Welsh-medium schools acquire English past tense and agreement morphology from an early school age, when the children are 4-years-old, into older primary school age, when they are 9-years-old, in a cross-sectional study design. Welsh-medium schools are bilingual schools where Welsh is the sole medium of instruction until Year 3 (age 7), when English is partly and gradually introduced to the curriculum.

The type of schooling and how it may impact the development of distinct language areas, in this case grammatical morphology, becomes particularly important for individuals

with language difficulties, such as for children with Specific Language Impairment (SLI). These children have been shown to have aggravated problems with the acquisition of inflectional morphology (Bishop, 2014; Leonard, 2014; Rice & Wexler, 1996) and may require greater educational support across the two languages (Ebbels, 2014). A number of studies with English-speaking children with SLI have reported that tense morphology constitutes a reliable clinical marker that discriminates between affected and unaffected children in both monolingual (Leonard, 1998; Leonard et al., 2007; Paradis & Crago, 2000) and bilingual contexts. In the latter case, the bilingual children examined are usually exposed to the two languages from birth or during the first three years of life and attend English mainstream preschool or primary schools in countries where English is the dominant language of the school and the community (Blom & Paradis, 2013; Paradis, 2010; Paradis, Schneider, & Duncan, 2013). Difficulties with tense morphology have also been documented for older school-age children attending Spanish-English bilingual classes in the US (Gutiérrez-Clellen, Simon-Cereijido, & Wagner, 2008; Jacobson & Schwartz, 2005). However, little is known about how these structures develop in contexts where English is the minority language both in the school and the community, but it may still be the language of treatment and diagnosis, as in the case of the Welsh-English bilingual context. In the present study, we investigate this question by examining tense and agreement acquisition in early school-age (4-to 6-years old) Welsh-English bilingual children attending Welsh-medium schools in North Wales. The investigation of these structures in young Welsh-English bilingual children will allow us to establish whether the differences between younger L2-TLD and L2-SLI children reported in previous studies where English was the medium of instruction (Paradis et al., 2017) also emerge in the present study where English is the minority language in the curriculum and the immediate community. We also examined how performance may be modulated by individual, child-internal (e.g. proficiency and non-verbal abilities) and language-specific properties (frequency, morphophonology) in line with previous studies on this issue (Blom & Paradis, 2013; Chondrogianni & Marinis, 2011; Paradis, 2011).

Tense in English and Welsh

Welsh is an inflectionally rich language with an intricate tense and agreement system (King, 2015). Present tense in Welsh is formed using a verbal compound consisting of an inflected auxiliary that carries tense and agreement features and the infinitival form of the lexical verb, as in (1a). Past tense can be formed in two ways: (i) periphrastically, by forming a verbal compound that combines inflected forms of the verb *bod* 'to be' or *gwneud* 'to do' with a bare lexical verb (1b), and (ii) synthetically, by adding the tense and agreement suffix to the stem (1c).

(1) a. **Mae** 'r hogyn **yn** gweld y clown.
Be.PRES.3SG the boy **ASP** see.INF the clown
'The boy sees/is seeing the clown.'

b. **Naeth.AUX** yr hogyn gweld y clown.
Do PAST.3SG 'The boy see.INF the clown.'

c. Gwel**-odd** yr hogyn y clown.
Saw.**PAST.3SG** the boy the clown
'The boy saw the clown'

Most lexical verbs in Welsh follow the regular rule formation of the synthetic past, apart from a small set of high frequency, irregular verbs, whose stem undergoes suppletion prior to suffix addition.

In English, agreement and regular past tense are marked through suffixal inflection on the verb stem. Agreement is only overtly marked in the present third person singular through the addition of *–s*, as in (2a), encoding the features of [-past], [3p], and [+sing]. Regular past

tense is consistently marked across all persons and numbers through the *–ed* suffix (2b). Irregular past tense is formed through suppletion, as in *ride → rode* (2c).

(2) a. John plays the guitar.

b. Betty painted the wall.

c. Garrett rode a bike.

d. He is playing guitar.

English also allows for the periphrastic formation of tenses through the use of auxiliaries, as in (2d).

In the following section, we review the acquisition of tense and agreement marking in English L2-TLD and L2-SLI children and its implications for the Welsh-English bilingual children in our study.

Acquisition of tense in English-speaking younger and older L2-TLD children:

Accuracy and error types

Existing studies with English-speaking L2-TLD children have examined the acquisition of tense and agreement morphology in two separate groups: either in younger preschool and early school-age children (Blom & Paradis, 2013; Blom, Paradis, & Duncan, 2012) or in older school-age children (Chondrogianni & Marinis, 2012; Chondrogianni & Marinis, 2011; Marinis & Chondrogianni, 2010; Paradis, Tulpar, & Arppe, 2016). Starting with agreement, in a longitudinal study on naturalistic data with 14 L2-TLD with different L1s, Blom et al. (2012) showed that L2-TLD children with a mean age of 5;6 years and with approximately two years of exposure to English produced 3SG *–s* at a rate of less than 20%. This rate increased to approximately 80% after one more year of exposure, although the increase was much starker for the children with inflecting L1s (approx. 90% accuracy) compared to children with non-

inflecting L1s (approx. 60% accuracy). Children with inflecting L1s had reached ceiling accuracy by the age of 7;6 years, whereas the children with non-inflecting L1s failed to do so (mean accuracy at the same age: 77%). For the L1 children with non-inflecting L1s, the development of 3SG *–s* continued well into older primary age. In a study by Paradis et al. (2016), Chinese L1-English L2 children aged between 8;5 to 10;5 years old, 78% of the children reached monolingual norms by the last round and after six years of exposure. This contrasts with the Turkish L1-English L2 children in the Marinis & Chondrogianni (2010) study who, by the age of nine years, 86% of them had reached age-appropriate norms.

Turning to the acquisition of regular and irregular past tense, studies with younger English L2-TLD children using the Test of Early Grammatical Impairment (TEGI, Rice & Wexler, 2001) have shown that their accuracy on regular verbs at the mean age of 5;6 years is approximately 50%, whereas accuracy on irregular verbs is even lower at 24% (Blom & Paradis, 2013). Omissions are the predominant error type for both regular and irregular verbs (approx. 50% and 43% for regulars and irregulars respectively), whereas overregularisations are secondary error types at this age (24%). This contrasts with the high proportion of overregularisations (approx. 80%) reported in Marinis & Chondrogianni (2010) study with older Turkish L1-English L2 children. Despite the fact that overall accuracy on irregular verbs remained low (approx. 20%), accuracy with regular verbs was almost at ceiling (approx. 90%). Although these studies seem to suggest that accuracy on regular verbs increases with age and that error types with irregular verbs change over time, they are not directly comparable given that they were carried out in different educational and social contexts (Canada vs. the UK) and included children with either a single L1 (Marinis & Chondrogianni, 2010) or with multiple L1s (Blom & Paradis, 2013). The present study aims to fill this gap by examining the development of these morphemes in a homogeneous group of Welsh-English younger and older L2-TLD children attending the same Welsh-medium schools. The present study diverges from previous studies in that the type of exposure that the children experience differs. Children attending Reception to Year 2 in Welsh-medium schools (4-6-year-old children), like the young

L2-TLD groups in our study, did not have any systematic exposure to English at the time of testing, although English was spoken within the wider school context and children had exposure to English through other extra-curricular literacy activities and the media (see Methods section). It is therefore expected that their accuracy may be overall reduced compared to the accuracy reported in other studies with English L2 children.

Acquisition of tense in English-speaking L2-SLI children: Accuracy and error types

A number of studies has shown that English-speaking L2-SLI children have exceptional problems with tense morphology even compared to their L2-TLD peers (Blom & Paradis, 2013; Gutiérrez-Clellen et al., 2008; Jacobson & Schwartz, 2005; Paradis, 2010), similarly to what has been reported for monolingual children with SLI (Bishop, 2014; Rice & Wexler, 1996). These differences are more pronounced during the early (pre)school years, when the L2-SLI children are between 4 and 8 years old (Blom & Paradis, 2013), and tend to fade away with age (Paradis, Jia & Arppe, 2017). In terms of inflectional morphemes, 3SG *–s* seems to be more affected compared to past tense marking (Gutiérrez-Clellen et al., 2008; Chondrogianni & Marinis, 2012) due to the complexity of semantic and syntactic information that 3SG *–s* carries. Studies comparing regular and irregular past tense marking in L2 children with and without SLI have reported conflicting results. Jacobson & Schwartz (2005) found that 8-year-old L2-TLD children were more accurate at producing regular verbs than irregular verbs, whereas L2-SLI children showed the reverse response. In contrast, studies by Paradis and colleagues (Blom & Paradis, 2013; Paradis, 2008) have found that both groups were more accurate with regular verbs than with irregular verbs and children with SLI performed worse than their TLD peers.

Differences between L2-TLD and L2-SLI children emerge not only with respect to accuracy but also with respect to the error types that the children produce. L2-SLI children have been shown to make fewer productive errors — that is, they are less likely to add

concatenating morphology to the verb stem in both correct and incorrect contexts, such as in the case of overregularisations with irregular verbs, and tend to omit inflectional morphemes more with both 3SG *–s* and regular past tense than their L2-TLD children. This finding is in line with research in both L1 (Leonard, McGregor, & Allen, 1992; Marchman, Wulfeck, & Weismer, 1999; Redmond & Rice, 2001) and L2 acquisition (Blom & Paradis, 2013; Jacobson & Schwartz, 2005).

Currently, there are no studies investigating how English tense and agreement develop in Welsh-English bilingual children with and without language impairment although both are official languages in Wales, they are taught in schools and used as official languages of treatment and diagnosis for children with language disorders. Research on the acquisition of verbal morphology in Welsh is also sparse, and, to our knowledge, there are only two studies to date investigating how verbal morphology is acquired in Welsh-speaking children.

Borsley & Jones (2001) examined the production of early verbal clauses in seven Welsh-speaking children aged between 1;6 and 2;5 years and reported that the earliest clausal utterances in the corpus did not contain finite verbs. Finite clauses with the suppletive forms of the copula bod *'to be'* were the first ones to emerge and were produced alongside non-finite clauses until the age of 2;5 years.

To date, the study by Chondrogianni & John (2018) is the only one to experimentally investigate how tense develops in the L1 Welsh of the same young Welsh-English bilingual children with TLD and SLI also reported here. The authors reported that Welsh-speaking L1-English L2-TLD children had almost ceiling accuracy in the periphrastic form of the past tense and were successfully prompted to produce the less frequent synthetic form of the past tense. L2-SLI children, on the other hand, were less likely to produce the synthetic form of the past tense following a prompt, compared to their age-matched L2-TLD peers with similar exposure and educational experience. This study concluded that early school-aged Welsh-speaking children with SLI are less likely to produce concatenating morphology on verbs in their L1.

The present study aims to investigate whether the productivity problems found in the children's L1 Welsh also feed into the acquisition of the L2 English. It also addresses whether English inflectional morphology is equally problematic for young Welsh-speaking children with SLI as in other studies with English-speaking L2 children. The case of Welsh-English bilingual children is noteworthy for both educational and linguistic reasons. The L2-SLI children in our study have not had any systematic exposure to English within the school system. As such, and given that children with language disorders may benefit more from instructed learning (Ebbels, 2014), the Welsh-English L2-SLI children may perform worse than the L2-SLI children in other studies. From a linguistic viewpoint, Welsh does not mark agreement overtly on present tense lexical verbs, it has a very small set of irregular verbs, and regular past tense formation on lexical verbs is productive, yet problematic for Welsh-speaking children with SLI (Chondrogianni & John, 2018).

Factors predicting tense acquisition in a usage-based approach

Various studies have revealed that accuracy and error patterns in tense marking are influenced by child-external factors, such as children's age and length of systematic L2 exposure, by child-internal factors, such as their first language and vocabulary size, as well as by more language-level and item-specific factors, such as word frequency and phonological properties of verbs (Blom & Paradis, 2013; Blom et al., 2012; Marinis & Chondrogianni, 2010; Paradis, 2005). Given that the population in the present study was rather homogeneous in terms of L1 background (Welsh), age and length of exposure to English (Welsh-medium (pre-)school education), we focused predominantly on the child-internal factors (vocabulary size, vocabulary skills, non-verbal intelligence/analytic reasoning) and language/item-specific factors (frequency, morphophonology) that modulate performance on English inflectional morphology.

Starting with child-internal factors, research with simultaneous bilingual and L2 children has shown an association between vocabulary size and grammatical development

(Conboy & Thal, 2006; Marchman, Martinez-Sussmann, & Dale, 2004), even when specific morphological properties, such as 3SG *–s* and past tense, are considered (Blom & Paradis, 2013; Chondrogianni & Marinis, 2011, Marinis & Chondrogianni, 2010). This holds for both typically developing and language-impaired children (Blom & Paradis, 2013). A second child-internal factor predicting L2 language outcomes in typically developing (Genesee & Hamayan, 1980; Masoura & Gathercole, 1999) and language-impaired children (Blom & Paradis, 2013) is analytic reasoning and pattern recognition skills measured though fluid, non-verbal intelligence, and it has even emerged as a significant predictor for L2 children's performance on tense morphology (Blom & Paradis, 2013; Paradis, 2011).

Turning to language-level factors, word, lemma and type frequency and the morpho-phonological properties of the verb stem have been argued to predict accuracy and error types with tense and agreement within usage-based accounts of language acquisition, such as Bybee's network model (Bybee, 2010). Word frequency refers to the frequency with which a particular verb occurs with a specific morpheme, e.g. *runs* or *kicked*. Lemma frequency refers to the frequency of a word regardless of its inflectional form, and type frequency to the number of lexemes with which an affix or its allomorphs occur, e.g. *kick-ed*, *play-ed*,. A schema with a high type frequency, such as those for the regular past tense allomorphs /t/ and /d/, is productive and is likely to apply to new words. A strong schema for regular past tense marking will lead to temporary overregularisations (e.g. *catched*) until the correct lexical representation for the irregular verb (e.g. *caught*) is sufficiently strong. Productivity is moreover a function of the degree of similarity between the words in a schema: The greater the variability of the schema, the more productive the schema will be (Blom & Paradis, 2013). The allomorph /Id/ is used with verbs that end in an alveolar stop (e.g. *wait, decide*); /t/ is used with verbs that end in a voiceless consonant (e.g. *stop, dance*); and /d/, constituting the largest class, is used with all other regular verbs. Following the logic of Bybee's model, each allomorph will have a separate schema with its own strength based on word frequency, type frequency, and variability, but the allomorphs will be linked together by semantics, that is, the past tense

meaning. Many irregular past tense forms in English are isolated, for example, *say–said* (Bybee, 2010). Irregular past tense acquisition in English is thus largely dependent on word frequency, in contrast to regular past tense acquisition, which relies on word frequency, type frequency, and variability. The allomorph /Id/, as the least productive of all three regular past tense allomorphs, is expected to rely more on word frequency than the other allomorphs and may contrast with the acquisition of /d/, which is the most productive allomorph class. Word frequency may influence children's accurate use of both irregular and regular forms. At the same time, morpho-phonological properties of the verb stem may influence productivity with both regular and irregular past tense formation, as this is another surface property that children may attend to (Bybee & Slobin, 1982).

Although studies have reported effects of frequency and morphophonology for monolingual children with and without SLI (Marchman et al., 1999; van der Lely & Ullman, 2001), only a handful of studies have investigated these factors in L2 children's acquisition of English tense and agreement inflections (Blom & Paradis, 2013; Marinis & Chondrogianni, 2010; Paradis et al., 2016, 2017) with conflicting results. Blom et al. (2012) reported an effect of word and lemma frequency in 5- to 7-year-old L2-TLD children's production of 3SG *–s*. This effect was more pronounced for children with non-inflecting L1s compared to the children with inflecting L1s. The authors explain this difference in two ways. First, children with inflecting L1s may have already built the necessary schemata for agreement or past tense formation in their L1 and this may transfer in the L2 schema formation based on overlap. Second, children with non-inflecting L1s may initially not attend to or perceive features such as person and number on verbs, as these associations are not present in their L1. Similar interactions were found between children's L1 and allomorph type. In the same study by Blom et al. (2012), accuracy on 3SG *–s* was almost at ceiling regardless of allomorph type for children with inflecting L1s.

Effects of frequency and morphophonology seem to interact with age. In the study by Blom & Paradis (2013) with 5;6-year-old L2-TLD and L2-SLI children, there were effects of

word frequency and allomorph type on regular past tense verbs and of word and lemma frequency on irregular verbs. Allomorph type effects were observed with overregularisations with verbs ending in an alveolar (/t/ or /d/) being overregularised less frequently compared to verbs with other endings. However, in the study by Paradis et al. (2017) with older (range from 7;8 to 9;8 years) L2-TLD and SLI children with both inflecting and non-inflecting L1s, word frequency effects were only reported for the irregular past tense. No effects of morphophonology or frequency were found for 3SG *–s* and regular past tense. Effects of word frequency only were also reported for the 7- to 9-year-old Turkish-speaking children in the Marinis & Chondrogianni (2010) study. These findings contrast with the effects of morphophonology that were reported for the 8- to 10-year-old Chinese-speaking children in Paradis et al. (2016) for 3SG *–s* and regular past. Taken together, these results suggest that younger children especially with non-inflecting L1s, tend to be more affected by word frequency and morphophonology. For children with non-inflecting L1s, these effects persist even at later stages of development, whereas for children with inflecting L1s, these effects are less likely to surface for 3SG *–s*, regardless of age, and are more likely to emerge in the context of past tense acquisition.

Effects of frequency and morphophonology may also be modulated by the impairment status. Given that children with SLI are less efficient at processing linguistic input compared to their unimpaired counterparts, they may be less efficient at making use of type frequency information and will instead rely on word frequency (Blom & Paradis, 2013). Insufficient parsing of the input may lead to incomplete schematisation and impact SLI children's productivity with both regular and irregular (overregularisations) past tense forms. The combination of reliance on word frequency along with lack of productivity may explain why SLI children may differ less from their TLD counterparts on irregular compared to regular past tense verbs. The L2-SLI children have been shown to be affected by word and lemma frequency in a similar way to their L2-TLD peers (Blom & Paradis, 2013; Paradis et al. 2016). However, L2-SLI children have been shown to overregularise less than their TLD peers and

to exhibit an opposite-than-expected pattern with past tense allomorphs, suggesting that they are less sensitive to surface-level morphophonological information compared to their TLD peers (Blom & Paradis, 2013).

In the present study, the Welsh L1-English L2-TLD children may transfer the schema of the synthetic past formation from Welsh onto regular past tense formation in English. The presence of suffixal inflectional morphology in the L1 verbal paradigm may place them on a par with children in other studies with inflecting L1s, who have also been found to be less sensitive to surface properties such as type frequency and morphophonology (Blom et al., 2012). However, these effects may be modulated by age and impairment. Any frequency effects may be more pronounced for the younger L2-TLD/SLI children. Children with SLI may additionally be less sensitive to surface properties due to difficulties with sufficient parsing of the linguistic input (Blom & Paradis, 2013).

Present study

The present study examined whether school-age Welsh-English bilingual children with typical development and Welsh-English bilingual children with language impairment have problems with the production of 3SG *–s* and past tense in English. The motivation for the study was threefold. First, we wanted to examine developmental changes in tense and agreement production from younger to older L2-TLD children with a homogeneous L1 (Welsh). Given that Welsh is an inflectionally rich language, the developmental pattern that we observe in this group of English L2 learners may differ from what is reported for Chinese L1 children (Paradis et al., 2016) and their performance may approximate more that reported for the Turkish L1-English L2 children by Marinis & Chondrogianni (2010), with the caveat that the L2 children in our study had reduced L2 exposure compared to previous studies. Second, we wanted to establish whether tense can be equally problematic for young Welsh-English bilingual children with language impairment attending Welsh-medium schools as it has been reported for other

English L2-SLI populations (e.g. Blom & Paradis, 2013). The population in the present study differs from other young L2-TLD/SLI children examined in previous studies (Blom & Paradis, 2013, 2015; Jacobson & Schwartz, 2005; Paradis et al. 2013) in that English was not the primary language of instruction and this may affect the children's proficiency (see also Guiterrez-Clellen et al., 2008 for Spanish-English bilingual children in bilingual schools). Third, we wanted to investigate how child-internal and language/item-specific factors modulate performance in L2-TLD and L2-SLI children. More specifically, we addressed the following research questions:

1. Do young L2-TLD children differ from older L2-TLD children on 3SG *-s* and regular and irregular past tense usage (accuracy and error types)?
2. Do young children with SLI differ from their age-matched L2-TLD controls on 3SG *-s* and regular and irregular past tense usage (accuracy and error types)?
3. How do vocabulary size and vocabulary skills, non-verbal skills, frequency, and phonological properties modulate accuracy in Welsh L1-English L2 children with typical development and with language impairment?

Method

Participants

Fifty-two children participated in the study (Table 1). There was a group of 25 Welsh-English 7-to 9-year-old L2-TLD children (L2-TLD_O) attending Years 3 and 4 in Welsh-medium schools, a younger group of 4-to 6-year-old L2-TLD children (*N*=17) from Reception to Year 2 in the same schools (L2-TLD_Y), and a group of young 4-to 6-year-old L2-SLI_Y children (*N*=10) attending the same schools as the TLD children in Bangor and the surrounding area (within a 20-mile radius around Bangor). The age difference between the L2-TLD_O and L2-TLD_Y groups was significant ($F(1,40)=93.19$, $p<.001$), whereas the L2-TLD_Y children were

age-matched with the L2-SLI_Y group (L2-TLD_Y: *M*=5;5, range: 54-76 months; L2-SLI_Y: *M*=5;3, range: 54-74) (*F(*1, 25)=4.12, *p*>.1)[1].

Welsh-medium schools are Welsh-English bilingual schools where Welsh is the main language of instruction until Year 3, when English is introduced and gradually takes up to 30% of instruction by key stage 2 (age 7-11) (Welsh Assembly Government, 2010). At the time of testing, the older L2 children had on average three hours of exposure to English per week. The data were collected as part of two independent projects entitled "Clinical markers of language impairment in Welsh-English bilingual children" and "Grammatical abilities of Welsh-English bilingual children". Ethics approval for the projects was obtained from Bangor University and the University of Edinburgh. L2 children were included in the L2-TLD group if their parents and/or teachers reported no concerns regarding their language development and if they did not meet the inclusion criteria for SLI (see below). From the L2-TLD group, children were included in the older L2-TLD group if they attended Years 3 or 4 in Welsh-medium schools, whereas they were included in the younger L2-TLD group if they were attending Reception and up to Year 2 in the same schools. None of the children across the three groups had any history of hearing impairment, attention deficit disorder, autism spectrum disorder, acquired neurological damage, or cognitive deficits. L2 children were included in the language impaired group if they met the following inclusion criteria (see Chondrogianni & John, 2018, for a detailed description of the profile of the L2-SLI group): (i) a formal diagnosis by a professional speech-language therapist (SLT); seven children were diagnosed as having SLI and were attending special classes at the time of testing, and (ii) a parent/teacher reported history of SLI (all children included in the bi-SLI group were reported to be late talkers), and/or

[1] We did not include language-matched children to the SLI children for both theoretical and practical reasons. Given the paucity of Welsh language measures, we would need to potentially rely on MLU measures, and the duration of the project did not allow us to gather long enough naturalistic data to make these calculations. We also believe that for educational and diagnostic purposes having age-matched children is more relevant.

there were concerns expressed by the parents or a professional school teacher about the child's language development, including speaking and understanding. Children were also classified as being at risk for language impairment if they had low language abilities (-1.25*SD* below the younger L2 group mean) across at least two language domains (phonology, vocabulary, expressive and receptive grammar) in both languages (L1/Welsh and L2/English) using the definition of bi-SLI as low language abilities across the two languages, not just one (Armon-Lotem & Meir, 2016). This resulted in three further children from the younger group of L2 children being included in the L2-SLI group, even though they did not have a formal diagnosis.

INSERT TABLE 1 ABOUT HERE.

Materials

British Picture Vocabulary Scale III (BPVSIII) (Dunn, Dunn, Styles, & Sewell, 2009). To assess children's vocabulary size and whether or not they have reached age-appropriate vocabulary skills in English we used the BPVSIII (Dunn et al., 2009). This is a comprehension of a single word vocabulary task standardised with monolingual British English-speaking children. In this task, children see a picture panel with four pictures and they are asked to point to the picture that goes with the word produced by the experimenter. The raw scores from the BPVS were converted into standard scores, as well as z-scores (Table 1). In the analysis, we took standard scores (BPVS_SS) to indicate differences in vocabulary abilities. Raw scores (BPVS_RS) were taken to be a proxy for vocabulary *size*, and thus, were analysed separately from the BPVS_SS to investigate how differences in vocabulary size modulate performance across the three groups independently of vocabulary skills. The BPVS_SS in Table 1 showed that the L2-TLD_O group did not differ from the L2-TLD_Y counterparts and both groups had reached age-appropriate norms, while the L2-SLI_Y group scored significantly lower compared to their 2-TLD_Y peers ($F(1,25)=13.19$, $p<.01$). The BVPS_RS revealed that the

L2-TLD_O group had significantly larger vocabularies than the younger counterpart (F(1, 40)=24.11, p<.001), whereas the L2-SLI_Y group had significantly smaller vocabularies compared to their age-matched L2-TLD_Y peers (F(1,25)=13.19, p<.01).

Raven's Coloured Progressive Matrices (Raven, 2003). To assess children's non-verbal intelligence and analytic reasoning skills through pattern recognition, we used the Raven's Coloured progressive matrices (Raven, 2003). In this task, children see a picture of a pattern with a piece missing and they are asked to select one out of six pieces that completes it. The task consists of 3 sets with 33 items overall (11 per set). All TLD and language impaired children scored within the normal range on the Raven's. In the present study, and given the age difference between the L2-TLD_O and the L2-TLD_Y and L2-SLI/TLD_Y children, we used the standard scores as predictors of children's performance in the analysis to correct for age.

Welsh morphosyntax. To assess children's abilities on Welsh morphosyntax, we developed a novel sentence repetition task (SRT) targeting Welsh verbal and nominal inflections and complex structures (Chondrogianni, Davies & Thomas, 2013). This SRT was developed within the European COST-Action IS0804 'Language Impairment in a Multilingual Society: Linguistic Aspects and the Road to Assessment' and followed the SRT specifications laid out by Marinis and Armon-Lotem (2015). There were six items across seven structures (VSO sentences, subject–verb agreement, prepositions, possessives, tense, relative clauses, subject and object *wh*-questions) resulting in 42 items in total. For consistency and comparability with the COST Action scoring procedures, we followed the scoring protocol from the British Clinical Evaluation of Language Fundamentals 2-preschool (Wiig, Secord & Seme, 2006), but instead of adding up the scores per item, we averaged it across items (Table 1). The L2-SLI_Y children differed from their L2-TLD_Y counterparts (F(1,25)=24.3, p<.001), whereas the two TLD groups did not differ from each other.

Production of 3SG –s and Past Tense. To assess children's production of the 3SG *–s* and the past tense, we used the screening probes of the TEGI (Rice & Wexler, 2001). The TEGI screener consists of two parts. The first part comprises 10 test items targeting 3SG *–s* present tense forms; the second part targets regular (10 items) and irregular (8 items) past tense verbs. In the TEGI 3SG *–s* probe, children are shown pictures of professionals engaging in activities and are prompted as follows: "Here is a teacher. Tell me what she does". The felicitous answer is with a 3SG present tense verb form, e.g. "A teacher teaches". In the past tense probe, children are shown pairs of pictures indicating a sequence of events. In the first picture, an individual is engaging in an action and in the second picture, the action/event is completed. Participants are prompted to produce the past tense with the following probe: "Here the boy is raking. Now he is done. Tell me what he did".

To ensure that the production of the target tense morphemes was not confounded with general difficulties producing these morphemes, all children were first tested on the phonological probe of the TEGI, which measured the word-final phonemes /d/, /t/, /s/, and /z/—the morphological affixes of interest in this study. All groups of children passed the phonological probe, suggesting that asymmetric scores between the phonological probe and the tense production would be ascribed to problems that lie in children's difficulties with tense morphemes rather than to phonological problems.

Background questionnaire. To assess children's family history, exposure and language use, we administered the short version of the Parents of Bilingual Children Questionnaire (PABIQ) (Tuller, 2015). The questionnaire elicited information about the child's quantity of exposure to the two languages, developmental milestones, (family) history of learning, and language disorders. We also measured children's input richness through the frequency of book-reading activities, singing songs, and watching TV or listening to the radio (scoring ranged between 1-almost never to 4-always and was averaged across the different activities). All children came

from predominantly Welsh-speaking homes (quantity of exposure for L2-SLI_Y: mean: 92.5%, *SD:10*; L2-TLD_Y: mean: 86.9%, *SD: 25*; L2-TLD_O: mean: 85.3%, *SD*: 21). Parents also reported that their children engaged in more extracurricular activities in Welsh (L2-SLI_Y: mean=3.3, SD=0.9; L2-TLD_Y: mean=3.6, SD=0.9*;* L2-TLD_O: mean=3.9, SD=0.6) than in English (L2-SLI_Y: mean=2.5, SD=1.3; L2-TLD_Y: mean=2.7, SD=1.4*;* L2-TLD_O: mean=3.3, SD=1). This difference between the two languages was significant across the groups (p<.001).

Calculation of word/lemma frequency. Word frequencies were obtained from the CELEX lexical database (Baayen et al., 1993) and have been used in previous studies to predict L2 children's performance on past tense (Marinis & Chondrogianni, 2010). Frequency lists were created by including word and lemma log frequencies for each verb. Log frequencies were counted to control for a possible confounding effect of word and lemma frequency (Blom & Paradis, 2013; Pinker & Ullman, 2002). Lemma frequency is the total of all (inflected) forms (bare stem, third person singular, past tense, progressive, and past participle) of a given verb.

Coding and scoring. Scorable responses for the 3SG *–s* probe were target-like answers (e.g. *flies*, *plays*), omissions (e.g. *fly*, *play*), and double-marked forms (e.g. *flieses*, *playses*). According to the TEGI guidelines, any other verb tense (other than 3SG *–s,* e.g. *flying, do fly*) is considered to be an unscorable response. However, in our study, these were counted as scorable responses, and more precisely as errors, because the children produced a substantial number of these (L2-TLD_O=8.4%, L2-TLD_Y=19.41%, L2-SLI_Y=19%). By including these types of responses as scorable responses, we increased the number of items for the analysis and we could better capture the (error) profile of L2 children with SLI. The final scorable responses therefore included correct/incorrect responses and other tense responses. Responses to which the child did not respond at all were counted as unscorable

(L2-TLD_O=0.8%, L2-TLD_Y=6.47%, L2-SLI_Y=49%).[2] The rest of unscorable responses across the groups consisted of responses in Welsh and fragments with nouns (L2-TLD_O=0.4%, L2-TLD_Y=0.6%, L2-SLI_Y=12%). Verbs were also coded for each of the allomorph type, i.e. /z/, /s/, and /Iz/.

For regular past tense, scorable responses were target-like tense marking (e.g. *painted, brushed*) and uninflected (zero) marking (e.g. *paint, brush*). Any lexical verb forms other than regular past tense (e.g. *painting, brushes*) were also considered scorable and examined separately for error analyses following the same rationale as for the present tense (L2-TLD_O: 6.6%, L2-TLD_Y: 11.93%, L2-SLI_Y: 9.47%). We included all verb types produced by children, even if they were different from the ones prompted in the TEGI, following the TEGI recommendations. If a child used repetitive fixed-expressions, such as verbs produced without a subject (e.g. *finished, stopped*), up to three uses of this response were included as scorable, with any more than three noted as unscorable following the TEGI manual's scoring procedure (Rice & Wexler, 2001:17). No responses were observed in the younger-aged groups (L2-TLD_Y: 15.34%, L2-SLI_Y: 56.84%), as well as forms in Welsh production (L2-TLD_Y: 3.98%, L2-SLI_Y: 7.37%). Verbs were also coded for past tense allomorph type, i.e., /d/, /t/, and /Id/.

For irregular verbs, target-like marking (e.g. *gave*, *ate),* unmarked stems (e.g. *give*, *eat),* overregularisations (e.g. *gived*, *eated*) including the irregular stem with a regular past suffix (e.g. *gived*) or irregular past tense forms with a regular past suffix as in double-marked forms (e.g. *ated, maded*) were all considered scorable responses. Any tensed verbs (e.g. giving, eats) other than the prompted were also included in the scorable responses (L2-TLD_O: 10.6%, L2-TLD_Y: 29.23%, L2-SLI_Y: 25.88%). Unlike the TEGI scoring, in the present study, we counted overregularisations as error types. This scoring allowed us to

[2] In the Blom & Paradis (2013) L2 children with SLI also produced many unscorable responses (over 30%). Given the reduced exposure to English, the proportion of unscorable responses in the present study is even greater.

investigate between-group differences in terms of erroneous responses (i.e. overregularisations, unmarked stems, and other tense forms) for irregular past tense in detail, as well as to gauge the degree of productive (e.g. overregularisation) errors the children made. Unscorable responses such as no responses were observed across the three groups (L2-TLD_O: 1%, L2-TLD_Y: 16.92%, L2-SLI: 55.29%) and the L2-SLI_Y group produced 2.35% of forms in Welsh. We also coded verbs for the morphophonological properties of the verb stem, i.e. for ending in a /t/-/d/, as in *write*, vs. 'other', as in *catch*, as the latter category has been shown to give rise to more overregularisations with both monolingual (Bybee & Slobin, 1982) and bilingual (Blom & Paradis, 2013; Chondrogianni & Marinis, 2010) children.

Procedure. Children were tested in a quiet room in their schools or in their homes by a Welsh-English bilingual research assistant. The data presented in this paper form part of a larger study where children participated in two separate hourly sessions, one in Welsh and one in English, and were tested on a range of language tasks.

Results

Given the binomial nature of the accuracy results, we ran a generalized linear mixed logistic regression with the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R environment (R Core Team, 2013). The regression analyses investigated 3SG *–s*, and regular and irregular past tense separately to address (1) between-group differences in accuracy and error types with tense morphology, and (2) effects of child-internal, such as vocabulary size and vocabulary skills, non-verbal IQ, and language-specific factors, such as word and lemma frequency and phonological properties of verbs, on L2 children's tense and agreement marking. In the accuracy and overregularisation analyses, Group membership (L2-TLD_Y, L2-TLD_O, L2-SLI_Y), the standard scores from the Raven's were entered as child-level fixed effects, whereas word and lemma frequency and allomorph type were entered as item-level fixed effects. To better understand the contribution of vocabulary to children's performance,

we ran two separate series of models across the different morphemes: one set of models with BPVS_SS to correct for age and to investigate main effects and interactions between vocabulary skills and group differences on performance accuracy. The second series of models included the BPVS_RS to investigate the effect of vocabulary size on accuracy and overregularisations. Random factors were Child and Item, with a random intercept for each item, and with a random intercept and slope for each child. Children's responses were coded as "True=1" or "False=0", indicating whether the child produced a correct or an incorrect response, while coding for error types indicated whether the child gave a specific erroneous answer or not. Fixed-effect predictors with more than two levels were contrasted to a specified reference level via treatment-coding. The L2-TLD_Y group served as a reference level in all models for comparisons with the L2-TLD_O and the younger age-matched L2-SLI_Y group. Word frequency (WordFreq) was the log-transformed word frequency in the CELEX corpus and was found to be highly correlated with lemma frequency (LemmaFreq) across three tense morphemes (3SG *–s*: $r(37)=.94$, $p<.001$; Past regular: $r(75)=.97$, $p<.001$; Past irregular: $r(59)=.96$, $p<.001$). To control for collinearity between two predictors, a decorrelated word frequency predictor (WordFreq-Residuals) was therefore created by predicting the variation in word frequency by lemma frequency. After decorrelation, there was a significant weak correlation between WordFreq and WordFreq-Residuals for all three tense morphemes (3SG *–s*: $r(37)=.34$, $p<.05$; Past regular: $r(75)=.26$, $p<.05$; Past irregular: $r(59)=.28$, $p<.05$). Starting with a full model with all predictor variables, stepwise backward elimination via nested model comparisons was applied to obtain the optimal model with significant effects from the fixed factors and/or their interactions. All child-internal and language specific factors were entered into the logistic regression for accuracy and overregularisations. For the error analysis for 3SG *–s* and regular past tense, we only considered Group as a fixed effect. Diagnostic statistics for outliers, leverage and influence (Zhang, 2016) were used to identify deviations that could influence estimates of models. We then obtained p-values via the likelihood ratio test using

chi-square and calculated the C-statistics assessing goodness-of-fit for the optimal model. Models with C-value ranges above 0.80 or higher indicate a good fit (Chatterji & Hadi, 2006).

3SG –s Tense Marking: Accuracy.

Figure 1 presents children's accuracy on the 3SG *–s* probe.

INSERT FIGURE 1 ABOUT HERE.

For this probe, 441 data points (for scorable items) for the three groups were analysed. For allomorph type (*/**s**/, /z/, /iz/*; reference level in bold), verbs with a final voiceless consonant (e.g. *eat, make*) were coded for */s/* (*N*=19), verbs ending in a final voiced consonant or vowel (e.g. *clean, blow,*) for */z/* (*N*=14), and verb stems ending in a sibilant (e.g. *brush, catch*) for */Iz/* (*N*=6). There were no significantly influential outliers according to studentized residuals with Bonferroni (p<.01).

INSERT TABLE 2 ABOUT HERE.

To examine how vocabulary skills modulate group accuracy, we ran models with the BPVS_SS. The BPVS_SS (main effects/interactions) did not contribute to any of the group variance, and were thus excluded from the optimal model, which included a main effect of Group and an interaction between LemmaFreq and Group (Model 1, Table 2). The L2-TLD_O children had significantly higher accuracy than their younger TLD peers, who in turn had higher accuracy than the L2-SLI_Y children. The model with the interaction effects with LemmaFreq was preferred over the model without its interaction ($\chi^2(2)=10.36$, p<.01). The significant interaction between Group and LemmaFreq showed that L2-SLI_Y and the L2-TLD_Y children did not perform better with verbs with a high lemma frequency, whereas verb forms with a higher lemma frequency were more often produced correctly by the L2-TLD_O group. The

optimal model with BPVS_RS (Model 2, Table 2) contained a main effect of vocabulary and an interaction effect between Group and LemmaFreq, similar to the interaction found in the BPVS_SS model. The model with the raw BPVS scores was also better than the model with just LemmaFreq ($\chi^2(1)=4.29$, $p=.038$). The optimal model indicated that the L2-TLD_O group tended to differ from the L2-TLD_Y children and the L2-SLI_Y children performed less accurately than their L2-TLD_Y peers. Raven's, WordFreq-Residuals and Allomorph type did not show a main effect or interactions nor did their inclusion significantly increase the model fit and they were therefore excluded from the optimal model. Models 1 and 2 had a C-value of 0.92, indicating an excellent performance.

3SG –s Marking: Error types.

Figure 2 presents children's error types on the 3SG *–s* probe.

INSERT FIGURE 2 ABOUT HERE.

Children's responses for 3SG *–s* error types were omissions and other tense forms (double-marked and other tenses; see *Coding and scoring section*), coded as "Omission=1" and "Other tense use=0"; Omission was taken as the reference level. A model was fitted to total 212 data points. The summary of the regression results is shown in Table 3.

INSERT TABLE 3 ABOUT HERE.

The regression results showed that there were no significant differences between the older and younger TLD groups in the use of both omissions and other tense. The L2-SLI_Y children produced significantly more errors in other tense than omissions compared to their age-matched TLD peers. The likelihood ratio test for the model in comparison with an intercept-

only model confirmed that Group as a fixed predictor had a significant effect ($\chi^2(2)=9.58$, $p<.001$).

Regular Past Tense: Accuracy.

Figure 3 presents children's accuracy on the regular past tense probe.

INSERT FIGURE 3 ABOUT HERE.

The regular past tense probe yielded 371 data points. For Allomorph type (***/d/***, */t/, /Id/*; bold indicates the reference level), regular verbs ending in a vowel or voiced phoneme (e.g. *tie, climb*) were coded for */d/* (*N*= 31), a stem-final voiceless phoneme (e.g. *brush*, *jump*) for */t/* (*N*=28), and other verbs endings in an alveolar stop (e.g. *paint, lift*) for */Id/* (*N*=18).

INSERT TABLE 4 ABOUT HERE

When the BPVS_SS (main effects/interactions) were entered into the model to investigate the contribution of vocabulary skills, they did not contribute to any of the group variance, and were thus excluded from the optimal model, which included a main effect of Group and an interaction between WordFreq-Residuals and Group (Model 1, Table 4). The model with the interaction between Group and WordFreq-Residuals was also preferred over the model without it ($\chi^2(1)=10.31$, $p<.001$). The significant interaction between Group and WordFreq-Residuals revealed that the L2-TLD_O group performed surprisingly worse with regular verbs with a higher word frequency, whereas for the L2-TLD_Y and the L2-SLI_Y children performance on verbs with high frequency improved. The optimal model with BPVS_RS (Model 2, Table 4) included interactions between Group and BPVS and between Group and WordFreq-Residuals. In the case of BPVS_RS, these interactions revealed that the difference between the L2-TLD_Y and the L2-SLI_Y groups was modulated by vocabulary size. To better

understand the nature of the interactions between Group and BPVS_RS, we compared the baseline model with only the main effects of BPVS_RS with the model with the interaction. In the baseline model, there was a significant effect of vocabulary size (*E*=0.928, *SE*=0.342, *Z*=2.712, *p*=0.007) and the L2_TLD-Y children differed from their L2-SLI_Y peers (*E*=-1.985, *SE*=0.653, *Z*=-3.041, *p*=.002). However, the model with the interaction between the raw BPVS_RS scores and Group was preferred over the model without the interaction ($\chi^2(2)=9.098$, *p*=.011), and the significant difference between the two younger groups disappeared when the interaction with BPVS_RS was entered into the model (Table 4). The lack of interaction between the L2-SLI_Y group and BPVS_RS highlights that, similarly to the L2-TLD_Y group, SLI children's accuracy on the past tense increased as vocabulary size grew, and this concomitant increase in accuracy and vocabulary size in the two groups led to the disappearance of the group effect. The L2-TLD_O children tended to perform more accurately than the L2-TLD_Y children and this was found both in the model without (*E*=0.76, *SE*=0.43, *z*=1.78, *p*=.075) and with the interaction (Model 2, Table 4). However, their increased vocabulary size did not give rise to higher accuracy, as the significant interaction between the L2-TLD_O and BPVS_RS shows**.** Raven's, LemmaFreq and Allomorph type did not emerge as significant predictors in either optimal models. Both optimal models (Models 1 and 2) provided an excellent fit to the data with a C-value of 0.98.

Past regular: Error types.

INSERT FIGURE 4 ABOUT HERE.

Figure 4 presents children's error types on the regular tense probe. In the simple logistic model with Group as a fixed factor, children's erroneous responses were coded as "Uninflected stems=1" and "Other tense use=0", and errors of uninflected stems were taken as a baseline. The model was fitted to 103 data points, containing these two errors. To rule out influential observations of the model, regression diagnostic tests were run and identified outliers

(*p*=.164), but removing those outliers did not affect the model output. The details are summarized in Table 5.

INSERT TABLE 5 ABOUT HERE.

The results showed that there was no significant difference between the older and younger L2-TLD groups neither between the L2-SLI_Y and the L2-TLD_Y groups in error types. The likelihood test of two models, the models with and without Group as a factor, revealed that Group did not improve the model fit ($\chi^2(2)=0.34$, $p=.845$).

Irregular Past Tense: Accuracy

INSERT FIGURE 5 ABOUT HERE

Figure 5 presents children's accuracy on the irregular past tense probe of a model fitted on 333 observations. Phonology as a factor (***other****, /d/-/t/,* reference level in bold) contained irregular verbs that ended in other phonemes (e.g. *blow*, *give*), coded as '*other*' (*N*=41) and verb stems with alveolar-stop endings (e.g. *ride*, *eat*) as '*/d/-/t/*' (*N*=20). Diagnostic statistics for the model showed that there were no influential outliers that needed to be dropped from the model (*p* <.001).

INSERT TABLE 6 ABOUT HERE

The optimal model (Table 6) contained interactions between Group, LemmaFreq and non-verbal intelligence. The results showed that the L2-TLD_O children were more accurate on irregular verbs than their L2-TLD_Y counterparts, whereas the L2-SLI_Y and the L2-TLD_Y

di not differ. Overall, children showed improved accuracy on verbs with higher LemmaFreq, but, surprisingly, the L2-TLD_O children tended to perform less accurately on irregular verbs with higher lemma frequency (p=.076). The L2-TLD_O children with higher non-verbal IQ scores also tended to be more accurate than the L2-TLD_Y group (p=.086). BPVS_SS and BPVS_RS did not emerge as significant predictors and were excluded from the optimal models. The optimal model in Table 6 indicated an excellent model fit with a C value of 0.96.

Irregular Past Tense: Error types.

INSERT FIGURE 6 ABOUT HERE

To examine children's production of different error types with irregular verbs, we conducted a multinomial logistic regression modelling using the mnlogit package on the data presented in Figure 6 (Hasan, Zhiyu, & Mahani, 2016). A model fitted to 243 data points with Group as the fixed effect was analysed. Error types for irregular past tense were classified into **overregularisations**, uninflected stems (i.e. omissions), and other tense use (i.e. tense other than irregular tense)—the reference level is in bold. Regression diagnostics revealed that there were no influential outliers observed across all error types (p<.01).

For the L2-TLD_Y and the L2-SLI_Y groups, there were no differences between error types (unmarked stems vs. overregularisations: p=.31; other vs. overregularisations: p=.16), whereas the L2-TLD_O produced significantly more overregularisations than unmarked stems or other error types (p<.001 in both cases) compared to the L2-TLD_Y.

Subsequently, we investigated which factors influenced children's overregularisations given that this was the predominant error type, especially for the L2-TLD_O children. For this analysis, 151 data points of scorable overregularisations were modelled. Children's responses

were coded as “overregularised=1” or “non-overregularised=0”. Phonological properties of irregular verbs contained two phonological categories, ***others*** (*N*=41) and */d/-/t/* (*N*=20) (the reference level in bold). Studentized residuals with Bonferroni confirmed that there were no outliers that were influential on the model (p<.01).

INSERT TABLE 7 ABOUT HERE

BPVS_SS (main effects/interactions) did not contribute to any of the group variance, and were thus excluded from the optimal model, which included a main effect of Group and an interaction between Phonology and Group (Model 1, Table 7). The interaction between Group and Phonology revealed that the L2-TLD_O children overregularised with irregular verbs ending in /d/ or /t/ than with other phonemes more often than the L2-TLD_Y group, whereas the L2-TLD_Y and the L2-SLI_Y children were not affected by the phonological properties of the verb stem. The optimal model with BPVS_RS (Model 2, Table 7) contained the interactions between Group and BPVS_RS and Group and Phonology. The models with the interactions were preferred over the models with the main effects only (BPVS_RS: $\chi^2(2)=14.10$, p<.001; Phonology: $\chi^2(2)=5.021$, p=.08) revealing that group differences were modulated by differences in these variables and especially vocabulary size. For example, in the model with only the main effect of BPVS_RS, there was a significant difference between the L2-TLD_Y and the L2-SLI_Y (*E*=-2.181, *SE*=1.037, *Z*=-2.103, p= .036) and the L2-TLD_O children (*E*=1.331, *SE*=0.696, *Z*=1.913, p=0.056). When the BPVS_RS by Group interaction was entered into the model (Table 7), the Group effect between the L2-TLD_Y and the L2_SLI_Y disappeared, suggesting that group differences were carried by differences in the vocabulary size between the young TLD and SLI children. The difference between the L2-TLD-O and the L2-TLD_Y children remained. Moreover, there was an interaction between the L2-TLD-O and BPVS, as for the older children with a larger vocabulary size led to fewer overregularisations, whereas this was not the case for the two groups of younger children. All other fixed effects

(WordFreq-Residuals, LemmaFreq, Raven's) did not improve the final optimal model ($\chi^2(3)=1.07$, $p=.79$) and were thus excluded. The optimal models had a C-value of 0.98, showing an excellent fit.

Discussion

The present study investigated: (i) whether 3SG *–s* and past tense acquisition profiles differed between different age groups of L2-TLD children, (ii) how young L2-SLI children performed on tense morphemes compared to their TLD peers, (iii) whether children across the groups were affected by child-internal factors, such as vocabulary size and non-verbal intelligence, as well as item-specific factors, such as word, lemma, type frequency and morphophonology. These aims were addressed in a cross-sectional design, where younger L2-TLD children were compared to their older L2-TLD counterparts and to a group of younger, age-matched L2-SLI children.

Development of tense and agreement in L2-TLD children

Starting with the comparison between younger and older L2-TLD children, we found developmental effects within a homogeneous group of Welsh-speaking English L2 children. The older L2 children outperformed their younger counterparts on regular past tense formation and produced more errors with irregular verbs. Productivity with past tense increased as the children grew older, as reflected in the increased accuracy with regular past verbs and in the proportion of overregularisation errors that the older children produced. The older and the younger L2-TLD groups did not differ on accuracy with irregular verbs. but were differentiated by their error types with irregular verbs. The L2-TLD_O children produced significantly more overregularisations and fewer unmarked forms with irregulars compared to their L2-TLD_Y peers. This suggests that children's past tense formation schemas become more productive with age. This is in line with previous studies with older L2-TLD children that report higher

overregularisation rates (Marinis & Chondrogianni, 2010) compared to studies with younger L2-TLD children (Blom & Paradis, 2013) attending English mainstream education. The advantage of the present study is that this developmental pattern and differential error types were observed in a relatively homogeneous group of L2 children with the same L1.

Turning to 3SG *–s,* the findings in the present study are in line with other studies indicating that L2 children's production rates for 3SG *–s* and regular past tense improve with increased exposure to English (Ionin & Wexler, 2002; Paradis et al., 2008). The overall accuracy rate for the younger group in our study was similar to that reported for the young L2-TLD children of similar age in Blom et al. (2012) (Round 1: mean age: 67 months, accuracy: approx. 18%). The older children in our study had lower accuracy compared to that of the older children from inflecting and non-inflecting L1s in Blom et al. (2012) (Round 5: mean age 92 months, accuracy: approx. 95%). This difference between the accuracy rates in the two studies may be due to the type of exposure that the groups experience. For the Welsh-English bilingual children in our study, English was the less dominant language both within the immediate societal and the educational context as opposed to the L2 children attending English mainstream education in other studies.

Tense and agreement in young L2-SLI children

In the second research question, we addressed whether tense and agreement marking are problematic for L2-SLI children attending bilingual schools in North Wales. We found that the L2-SLI children performed more poorly than their age-matched L2-TLD peers on the two morphemes involving concatenating morphology, namely 3SG *–s* and regular past, when differences in vocabulary skills were controlled for through the use of the BPVS standard scores. Differences between the two groups were revealed both in terms of accuracy and error types. The L2-SLI children in our study omitted more in the context of 3SG *–s* and this gave rise to depressed accuracy, which contrasts with the results reported by Paradis et al. (2017) with older L2 children, where no group differences between the L2-TLD and L2-SLI children

were reported. These results suggest that problems with inflectional morphology may be more obvious when L2-SLI children are younger (Paradis et al, 2017). In the present study, the difference between the typically developing and the SLI children may be accentuated by the reduced length and the type of exposure. Given that children with SLI may be facilitated more by instructed exposure (Ebbels, 2014), lack of instruction in English may put L2-SLI children at a disadvantage in comparison to SLI children attending schools where both languages are more balanced in the curriculum.

Our study also revealed that children with SLI had difficulties with regular past tense formation. Differences between the two groups in regular past tense formation are well documented with other L2 child populations (Blom & Paradis, 2013; Guittierez-Clellen et al., 2008; Jacobson & Schwartz, 2005) and highlight the inherent problems with productivity that children with SLI have. This was further evidenced in the greater proportion of overregularisation errors that the L2-TLD_Y children made in comparison to their L2-SLI_Y counterparts, when differences in their vocabulary skills were controlled for through the use of the BPVS standard scores.[3] This is in line with other studies that have shown that L2-SLI children made fewer overregularisations than their TLD peers (Blom & Paradis, 2013; Jacobson & Schwartz, 2005), and despite the fact that the L2-SLI participants in the Jacobson & Schwartz (2005) study were 2 to 4 years older (*M*age=96 months) and had longer exposure to English (*M*exposure=6;9 years) than the L2-SLI children in our study. Overregularisations in the present study were observed despite the fact that the L2-TLD_Y children had received less intense and systematic English L2 exposure compared to the TLD children in the Blom & Paradis (2013) study. It could be hypothesised that the more intense and systematic English L2 exposure in the Canadian context may lead to the formation of stronger L2 schemas and

[3] Note that the differences between the L2-TLD_Y and the L2-SLI_Y groups were observed also in the baseline models with the BPVS_RS without the interaction. In the study by Blom & Paradis (2013) only the models without the interaction are reported and hence it is not clear, as is in the present study, whether or not group effects would be cancelled out when the effect of vocabulary in each group was considered.

higher productivity at least in the younger TLD group. At the same time, the lack of L2 English instruction may place the L2-SLI children in our study at a disadvantage compared to the SLI children in the Canadian study for reasons mentioned above. Importantly, the Welsh-English L2-SLI children in our study also had problems with concatenating morphology in their L1 (Chondrogianni & John, 2018). It seems, thus, that the problems with productive morphology that the L2-SLI_Y children face in their L1 transfer to their L2 (English) and may be accentuated by the lack of intense and lengthy L2 exposure. Taken together, these findings suggest that the trajectories of L2 tense acquisition reflect the complexity of modulating factors such as age, length and intensity of L2 exposure.

Impact of vocabulary size, non-verbal intelligence, word frequency, and phonology on tense marking

In the present study, general vocabulary size as measured through the BPVS_RS emerged as a significant child-internal predictor of children's performance on tense and agreement morphemes, and in the case of the younger TLD and SLI children, it explained most of the variance associated with the group differences. Importantly, vocabulary size effects were modulated by age. The two younger groups of L2-TLD and L2-SLI children were more accurate on 3SG *–s* and past tense, as their vocabulary increased, and similar effects were also found for the L2-TLD_O group in the case of 3SG *–s*. This is in line with previous studies with L2 children that have reported a relationship between children's vocabulary size and performance on inflectional morphemes within the same age group (Blom et al., 2012; Simon-Cereijido & Gutiérrez-Clellen, 2009). Interestingly older children overregularised less with irregular verbs as their vocabulary size increased, whereas for the younger L2-TLD and L2-SLI children, increased vocabulary also gave rise to more increased morphological productivity through overregularisations.

Non-verbal intelligence as measured through pattern recognition was also found to be a significant predictor of performance on inflectional morphology yet only for irregular verbs.

Previous studies reported an effect of non-verbal intelligence on the overall accuracy with tense morphology (regular and irregular past tense collapsed together) (Paradis, 2011) or on regular past but not on irregular past (Blom & Paradis, 2013). This finding suggests that general analytic skills may be an important source of individual differences in contexts where the L2 is the minority language similarly to previous studies with L1 minority children in an L2 majority context (Blom & Paradis, 2013; Genesee & Hamayan, 1980; Masoura & Gathercole, 1999; Paradis, 2011; Paradis & Jia, 2017), although how exactly they modulate specific grammatical structures merits further investigation.

Turning to item-specific variables, we did not find effects of morphophonology in the acquisition of regular past tense and 3SG *–s* in the L2-TLD_Y/O and L2-SLI children. Contrary to the L2 children with non-inflecting L1s in the Blom et al. (2012) study who were sensitive to type (allomorphic) frequency, the Welsh L1 – English L2 children in the present study did not exhibit such effects. Only the L2-TLD_O children exhibited some sensitivity to phonological properties when overregularising, although these effects showed the opposite than expected pattern. This finding may be due to the fact that the children in the present study speak an inflecting L1, Welsh, which has a highly regularised morphology. This is more in line with the performance of the L1 Turkish-speaking children in Marinis & Chondrogianni (2010) and contrasts with that of the L1 Chinese-speaking children in Paradis et al. (2016), confirming that strong productive L1 schemas make children less sensitive to surface-level L2 morphophonological properties and allow them to overregularise more freely.

The results from the effects of word and lemma frequency are mixed. Higher lemma frequency gave rise to higher accuracy with 3SG *–s* in the L2-TLD_O group. However, lemma frequency had a reverse effect on children's production of the irregular past tense, especially the L2-TLD_O and the L2-SLI_Y groups. This is a rather surprising finding that may be related to the type of corpus that we used for the frequency calculation in the present study. In the Blom et al. (2012) study, frequency effects emerged only in relation to the Edmonton English Language Learner corpus and not when the adult spoken corpus was considered (Blom et al.,

2012: 987). The fact that the frequency corpus used in our study was based primarily on written adult frequency data from the CELEX may have provided less representative information about the input received by the children in our study and may have masked any true frequency effects.

Conclusions and future directions

The present study examined the acquisition of tense and agreement morphology in three groups of Welsh-dominant English L2 children and investigated the child-internal and language-specific factors that affect performance on these morphemes. We found developmental effects both in terms of accuracy and error types in a cross-sectional study design with children from a homogeneous linguistic and educational background. The six- to nine-year-old L2-TLD_O children produced significantly more inflected regular past tense verbs and overregularisations with irregular verbs than their younger TLD peers and marginally more inflected 3SG verbs. These findings show that accuracy increases and error types change as a function of age even in a context where the L2 is the minority in the educational and societal context. However, a caveat is important here. Within the Welsh-medium school context, it is hard to disentangle effects of age from effects of instructed exposure to English, as the children grow older and attend higher school years. This is because the children attending Year 3 in our study were not only older than the children attending Years 1 and 2 but were also the ones receiving some systematic instruction in English. Although teaching of English was restricted to up to three hours a week in the schools we tested, it still remains a possibility that the age effects found in our study are modulated by English instruction. To more successfully disentangle effects of age from effects of instruction, future studies could explore how accuracy and error types with inflectional morphemes change in groups of Welsh-English bilingual children from different backgrounds, e.g. for simultaneous Welsh-English bilingual children or for English L1-Welsh- L2 children attending bilingual schools in Wales with different degrees of exposure to and instruction of both languages.

In terms of impairment effects, results revealed that four- to six-year-old Welsh-English bilingual children with SLI have pronounced problems with concatenating morphology, 3SG –*s* and regular past tense, in comparison to their L2-TLD peers, whereas the two groups have equally poor performance on irregular verbs. These findings confirm that productivity problems are present in children with SLI and distinguish them from their TLD peers. In the present study, the low performance of the L2-SLI children may be accentuated by the reduced L2 English input and lack of L2 instruction. L2-SLI children's low proficiency was also evidenced in the high rate of unscorable responses that they produced and in their low raw and standard vocabulary scores compared to their L2-TLD_Y peers, as well as compared to other L2-SLI children of similar age in previous studies (Blom & Paradis, 2013). Given that in this project we only included young children with SLI and the sample was rather small, future studies could explore whether these differences are found in a larger sample of children and persist for older Welsh-English bilingual children with SLI, as L2 exposure and systematic teaching increase.

Finally, in the present study, vocabulary size as indicated by the BPVS raw scores and lemma frequency were the major predictors of children's performance confirming the interplay between child-internal and language-specific factors and the development of inflectional morphology in the context of typical and impaired child L2 acquisition (Blom & Paradis, 2013). The lack of clear type frequency/allomorphic effects may be due to the fact that the Welsh L1 children spoke an inflecting L1 and were, hence, less sensitive to surface phonological L2 properties. Future studies where the item-level properties of the experimental material have been better controlled for than in the TEGI should investigate these effects in more detail.

References

Armon-Lotem, S., & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal of Language and Communication Disorders*, 51(6), 715-731. https://doi.org/10.1111/1460-6984.12242.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/https://doi.org/10.1016/j.jml.2007.12.005

Bishop, D. V. M. (2014). Problems with tense marking in children with specific language impairment: not how but when. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *369*(1634). https://doi.org/10.1098/rstb.2012.0401

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Consortium, C.-2, & Consortium, C.-2. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Consortium, C., & Consortium, C. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *Plos One*, *11*(7). https://doi.org/10.1371/journal.pone.0158753

Blom, E., & Paradis, J. (2013). Past tense production by English second language learners with and without language impairment. *Journal of Speech, Language and Hearing Research*, *56*(1), 281–294. https://doi.org/10.1044/1092-4388(2012/11-0112)

Blom, E., & Paradis, J. (2013). Past Tense Production by English Second Language Learners With and Without Language Impairment. *Journal of Speech, Language and Hearing Research*, *56*(1), 281. https://doi.org/10.1044/1092-4388(2012/11-0112)

Blom, E., Paradis, J., & Duncan, T. S. (2012). Effects of Input Properties, Vocabulary Size, and L1 on the Development of Third Person Singular –s in Child L2 English. *Language Learning*, *62*(3), 965–994. https://doi.org/10.1111/j.1467-9922.2012.00715.x

Borsley, R., D. & Jones, B. M. (2001). The development of finiteness in Early Welsh. *Journal of Celtic Language Learning*, 6, 9-20.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge University Press.

Chondrogianni, V., & John, N. (2018). Tense and plural formation in Welsh-English bilingual children with and without language impairment. *International Journal of Language & Communication Disorders*, *53*(3), 495-514. doi:10.1111/1460-6984.12363.

Chondrogianni, V., Davies, P. & Thomas, E. (2013). A new sentence repetition task for Welsh-speaking school aged children. Bangor: Bangor University.

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*(3), 318–345. https://doi.org/10.1075/lab.1.3.05cho

Chondrogianni, V., & Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism: Language and Cognition*, *15*(1), 5–21. https://doi.org/10.1017/S1366728911000368

Conboy, B. T., & Thal, D. J. (2006). Ties Between the Lexicon and Grammar: Cross-Sectional and Longitudinal Studies of Bilingual Toddlers. *Child Development*, *77*(3), 712–735. https://doi.org/10.1111/j.1467-8624.2006.00899.x

Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009). *The British Picture Vocabulary Scale III* (3rd ed.). London: GL Assessment.

Ebbels, S. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy*, *30*(1), 7–40. https://doi.org/10.1177/0265659013512321

Education and Lifelong Learning Committee. (2008). *Policy Review: Welsh language in education*.

Genesee, F., & Hamayan, E. V. (1980). Individual differences in second language learning. *Applied Psycholinguistics*, *1(1)*, 95–110.

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Wagner, C. (2008). Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Applied Psycholinguistics*, *29*(1), 3–19. https://doi.org/10.1017/S0142716408080016

Hasan, A., Zhiyu, W., & Mahani, A. S. (2015). *mnlogit: Multinomial Logit Model. R Package Version 1.2.4*. Retrieved from https://cran.r-project.org/web/packages/mnlogit

Jacobson, P. F., & Schwartz, R. G. (2005). English past tense use in bilingual children with language impairment. *American Journal of Speech and Language Pathology*, *14*(4), 313–323. https://doi.org/10.1044/1058-0360(2005/030).

King, G. (2015). *Modern Welsh: a comprehensive grammar*. London: Routledge.

Leonard, L. (2014). *Children with Specific Language Impairment* (2nd ed.). Cambridge, MA: MIT Press.

Leonard, L. B., McGregor, K. K., & Allen, G. D. (1992). Grammatical morphology and speech perception in children with specific language impairment. *Journal of Speech and Hearing Research*, *35*(5), 1076–1085. doi:10.1044/jshr.3505.1076.

Marchman, V. A., Martinez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: evidence from bilingual language learners. *Developmental Science*, *7*(2), 212–224. https://doi.org/10.1111/j.1467-7687.2004.00340.x

Marchman, V. A., Wulfeck, B., & Weismer, S. E. (1999). Morphological productivity in children with normal language and SLI: A study of the English past tense. *Journal of Speech Language and Hearing Research*, *42*(1), 206–219.

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*(1), 19–28. https://doi.org/10.3109/17549500903434125

Paradis, J. (2005). Grammatical morphology in children learning English as a second language: Implications of similarities with specific language impairment. *Language Speech and Hearing Services in Schools*, *36*(3), 172–187. https://doi.org/10.1044/0161-1461(2005/019)

Paradis, J. (2010). The interface between bilingual development and specific language

impairment. *Applied Psycholinguistics*, *31*(2), 227–252. https://doi.org/10.1017/S0142716409990373

Paradis, J. (2011). Individual differences in child English second language acquisition Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237. https://doi.org/10.1075/lab.1.3.01par.

Paradis, J. & Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors shape individual differences in catching up with monolinguals. *Developmental Science*, 20, e12433. DOI: 10.1111/desc.12433.

Paradis, J., Jia, R., & Arppe, A. (2017). The acquisition of tense morphology over time by English second language children with specific language impairment: Testing the cumulative effects hypothesis. *Applied Psycholinguistics*, *38*(4), 881–908. https://doi.org/10.1017/S0142716416000485

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language and Hearing Research*, *56*(3), 971–981. https://doi.org/10.1044/1092-4388(2012/12-0050)

Paradis, J., Tulpar, Y., & Arppe, A. (2016). Chinese L1 children's English L2 verb morphology over time: individual variation in long-term outcomes. *Journal of Child Language*, *43*(3), 553–580. https://doi.org/10.1017/S0305000915000562.

R Core Team. (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Raven, J. C. (2003). *Raven's Coloured Progressive Matrices^{TM} (CPM)*. London: Pearson Assessment.

Redmond, S. M., & Rice, M. L. (2001). Detection of irregular verb violations by children with and without SLI. *Journal of Speech, Language and Hearing Research*, *44*(3), 655–669. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11407569.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language

impairment in English-speaking children. *Journal of Speech and Hearing Research*, *39*(6), 1239–1257. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8959609

Rice, M. L., & Wexler, K. (2001). *Test of Early Grammatical Impairment*. NewYork: The Psychological Corporation.

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 301–330). London: Multilingual matters.

Van Der Lely, H. K. J., & Ullman, M. T. (2001). Past tense morphology in specifically language impaired and normally developing children. *Language and Cognitive Processes*, *16*(2–3), 177–217. https://doi.org/10.1080/01690960042000076

Welsh Assembly Government. (2010). *Welsh-medium Education Strategy*.

Wiig, E. H., Secord, W., & Seme, E. (2006). CELF-Preschool 2 UK. London: Pearson Education.

Zhang Z. (2016). Residuals and regression diagnostics: focusing on logistic regression. *Annals of Translational Medicine*, *4*(10), 195. https://doi.org/10.21037/atm.2016.03.36.

Table 1. Age at testing (in months) for Welsh-English L2 children and results from the BPVS-II, the Raven's and the Welsh SRT (Mean, Range and *SD*)

	Age	**BPVS**			**Raven's**			**Welsh SRT**
	(In months)	RS	SS	z-score	RS	SS	z-score	RS
L2-TLD_Y	67	66.12	88	-0.80	15.80	92.35	-0.17	2.2
(n=17)	54–76	67–137	63–106	-2.45–0.40	9-22	81–124	-1.25–1.60	1.4-2.7
	7	*16.11*	*11.5*	*0.77*	*3.9*	*13.61*	*0.9*	*0.3*
L2-TLD_O	93.72	91.56	82.08	-1.20	25.72	108.96	0.59	2.4
(n=25)	79–112	15–94	69–103	-2.1–0.2	16–33	81–124	-1.25–1.60	1.14-3
	9.83	17.03	9.87	0.66	4.85	10.88	0.72	1.2
L2-SLI_Y	61.3	42.5	77.6	-1.5	17.6	104.7	0.31	1.5
(n=10)	54–74	19–64	63–85	-2.45– -1.00	13–25	90–119	-0.65 –1.25	1.2-1.9
	7.13	14.98	7.06	0.46	3.84	12.30	0.81	0.2

Note. BPVS=British Picture Vocabulary Scale; Raven's=Raven's Coloured Progressive Matrices; L2=second language; TLD=typically developing; SLI=specific language impairment; SRT = Sentence repetition task (numbers indicate the CELF-type scoring, namely average out of a ceiling score of 3 per item per participant).

Table 2. Optimal logistic regression models for accuracy with 3SG *–s*.

Models	**Estimate**	***SE***	Z	**p**
Model 1				
Intercept	-3.853	0.400	-9.621	< .001 ***
L2-SLI_Y	-2.808	0.895	-3.136	.002 **
L2-TLD_O	1.238	0.366	3.379	< .001 ***
LemmaFreq	-0.021	0.316	-0.066	.95
L2-SLI_Y × LemmaFreq	-1.598	0.797	-2.006	.045 *
L2-TLD_O × LemmaFreq	0.496	0.250	1.982	.047 *
Model 2				
Intercept	-3.675	0.392	-9.377	< .001 ***
L2-SLI_Y	-2.33	0.913	-2.552	.011 *
L2-TLD_O	0.728	0.416	1.751	.078
BPVS_RS	0.499	0.243	2.060	.039 *
LemmaFreq	-0.020	0.316	-0.065	.95
L2-SLI_Y × LemmaFreq	-1.599	0.797	-2.007	.045 *
L2-TLD_O × LemmaFreq	0.495	0.250	1.978	.048 *

Note. Group: The L2-TLD_Y children taken as the reference level; L2-SLI_Y= young L2 children with SLI; L2-TLDO=older L2 typically developing children; BPVS_RS = raw scores on the British Picture Vocabulary Scale; LemmaFreq=log-transformed lemma frequency.

*p < .05. **p < .01. *** p <. 001.

Table 3. Logistic regression model for error types (omission vs. other tense forms) with 3SG *–s*.

	Estimate	***SE***	*Z*	***p***
Intercept	-2.325	0.749	-3.103	.002 **
L2-SLI_Y	2.756	1.090	2.528	.011 *
L2-TLD_O	-0.102	0.779	-0.131	.89

Note. Group: Group: The L2-TLD_Y children taken as the reference level; L2-SLI_Y= young L2 children with SLI; L2-TLDO=older L2 typically developing children; Other tense forms = double-marked present tense forms and other tenses)

*$p < .05$. ** $p <. 01$.

Table 4. Optimal logistic regression models for accuracy with regular past tense.

Models	Estimate	*SE*	*Z*	*p*
Model 1				
Intercept	-10.565	1.157	-9.135	< .001 ***
L2-SLI_Y	-2.532	0.629	-4.021	< .001 ***
L2-TLD_O	1.353	0.367	3.686	< .001 ***
WordFreq-Residuals	1.001	0.695	1.441	.15
L2-SLI_Y × WordFreq-Residuals	0.640	0.518	1.237	.22
L2-TLD_O × WordFreq-Residuals	-0.549	0.232	-2.366	.018 *
Model 2				
Intercept	-10.174	1.161	-8.766	< 0.001 ***
L2-SLI_Y	-0.579	1.03	-0.563	.57
L2-TLD_O	1.014	0.403	2.515	.012 *
BPVS_RS	1.265	0.458	2.762	.006 **
L2-SLI_Y × BPVS_RS	0.791	0.987	0.801	.423
L2-TLD_O × BPVS_RS	-1.283	0.550	-2.332	.02 *
WordFreq-Residuals	0.991	0.694	1.428	.153
L2-SLI_Y x WordFreq-Residuals	0.702	0.533	1.318	.188
L2-TLD_O × WordFreq-Residuals	-0.542	0.232	-2.342	.019 *

Note. Group: The L2-TLD_Y children taken as the reference level; L2-SLI_Y= young L2 children with SLI; L2-TLDO=older L2 typically developing children; BPVS_RS = raw scores on the British Picture Vocabulary Scale; WordFreq-Residuals= log-transformed word frequency created by the variation in word frequency by lemma frequency.

*p < .05. **p < .01. *** p <. 001.

Table 5. Logistic regression model for error types (omissions vs. other tense forms) with regular past tense.

	Estimate	***SE***	***Z***	***p***
Intercept	-0.40	0.585	-0.683	.49
L2-SLI_Y	0.534	0.934	0.572	.57
L2-TLD_O	0.055	0.725	0.076	.94

Note. Group: The L2-TLD_Y children taken as the reference level; L2-SLI_Y= young L2 children with SLI; L2-TLD_O=older L2 typically developing children; other tense forms = double-marked past tense forms and other tenses.

Table 6. Optimal logistic regression model for accuracy with irregular past tense.

	Estimate	***SE***	**Z**	**p**
Intercept	-7.685	1.02	-7.535	<.001 ***
L2-SLI_Y	-0.729	0.844	-0.864	.388
L2-TLD_O	1.137	0.58	1.961	.049 *
LemmaFreq	1.155	0.551	2.097	.036 *
L2-SLI_Y × LemmaFreq	-0.943	0.729	-1.294	.196
L2-TLD_O × LemmaFreq	-0.6567	0.371	-1.771	.077
NVIQ	-0.301	0.341	-0.882	.378
L2-SLI_Y × NVIQ	0.128	0.809	0.159	.874
L2-TLD_O × NVIQ	0.791	0.461	1.715	.086

Note. Group: Second language typically developing younger children (L2-TLD_Y) was taken as the reference level; L2-SLI_Y=second language impaired children; L2-TLD_O=second language typically-developing older children, LemmaFreq=log-transformed lemma frequency; NVIQ=standard scores on Raven's Progressive Matrices (non-verbal IQ).
*p < .05. *** p <. 001.

Table 7. Optimal logistic regression models for overregularisations with irregular verbs.

Models	Estimate	*SE*	*Z*	*p*
Model 1				
Intercept	-12.186	1.465	-8.321	< .001 ***
L2-SLI_Y	-2.532	0.976	-2.595	< .01 **
L2-TLD_O	1.728	0.589	2.935	< .01 **
Phonology (*/d/-/t/*)	-0.406	1.715	-0.237	.81
L2-SLI_Y × Phonology (*/d/-/t/*)	1.021	1.401	0.729	.47
L2-TLD_O × Phonology (*/d/-/t/*)	1.395	0.697	2.001	< .045 *
Model 2				
Intercept	-11.77	1.465	-8.03	< .001 ***
L2-SLI_Y	5.010	4.186	1.197	.231
L2-TLD_O	2.042	0.623	3.278	.001 **
BPVS_RS	1.32	0.721	1.829	.067
L2-SLI_Y × BPVS_RS	10.005	7.790	1.284	.199
L2-TLD_O × BPVS_RS	-2.317	0.856	-2.707	.007 **
Phonology (*/d/-/t/*)	-0.379	1.714	-0.221	.825
L2-SLI_Y × Phonology (*/d/-/t/*)	0.791	1.540	0.513	.608
L2-TLD_O × Phonology (*/d/-/t/*)	1.377	0.691	1.993	.046 *

Note. Group: The L2-TLD_Y children taken as the reference level; L2-SLI_Y= young L2 children with SLI; L2-TLD_O=older L2 typically developing children; Phonology=phonological properties of irregular verbs; *others*=verbs with a stem that ended in other sounds taken as the reference level; */d/-/t/* verbs with a stem final /d/ or /t/.
*p < .05. **p < .01. *** p <. 001.

Figure 1. Accuracy of 3SG –*s* for the L2-TLD_O, the L2-TLD_Y and the L2-SLI_Y children.

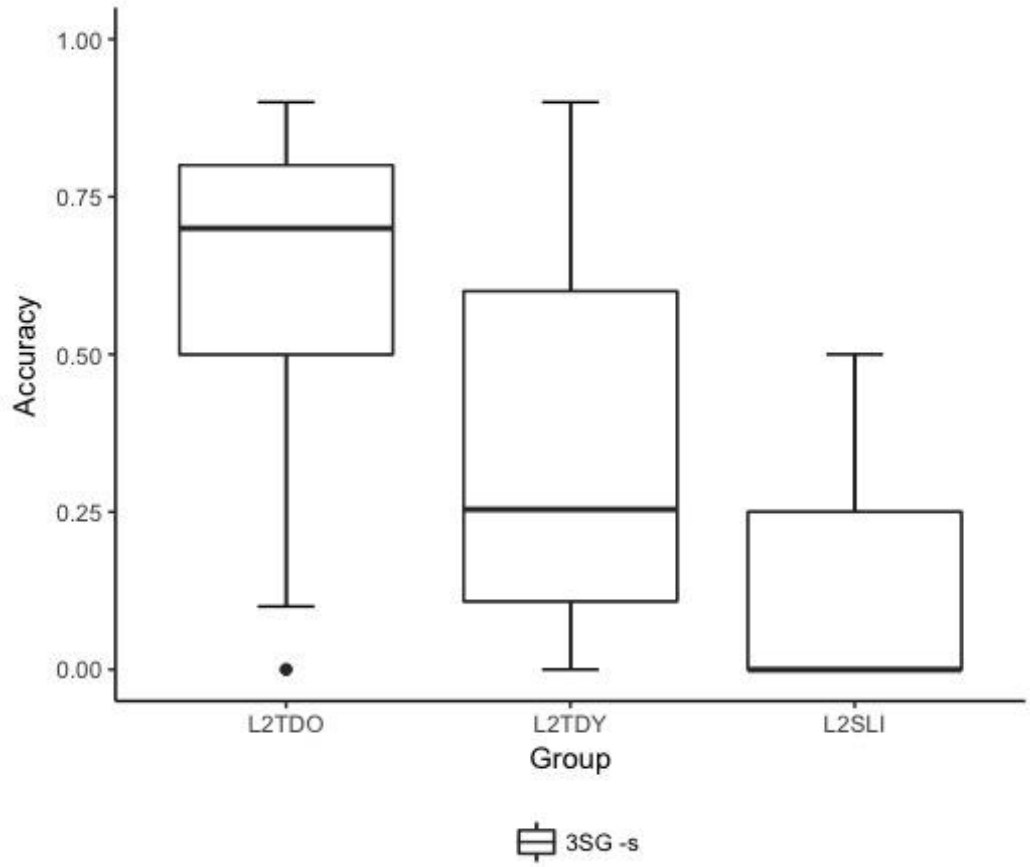


Figure 2. Error types of 3SG *–s* for the L2-TLD_O, the L2-TLD_Y and the L2-SLI_Y children.

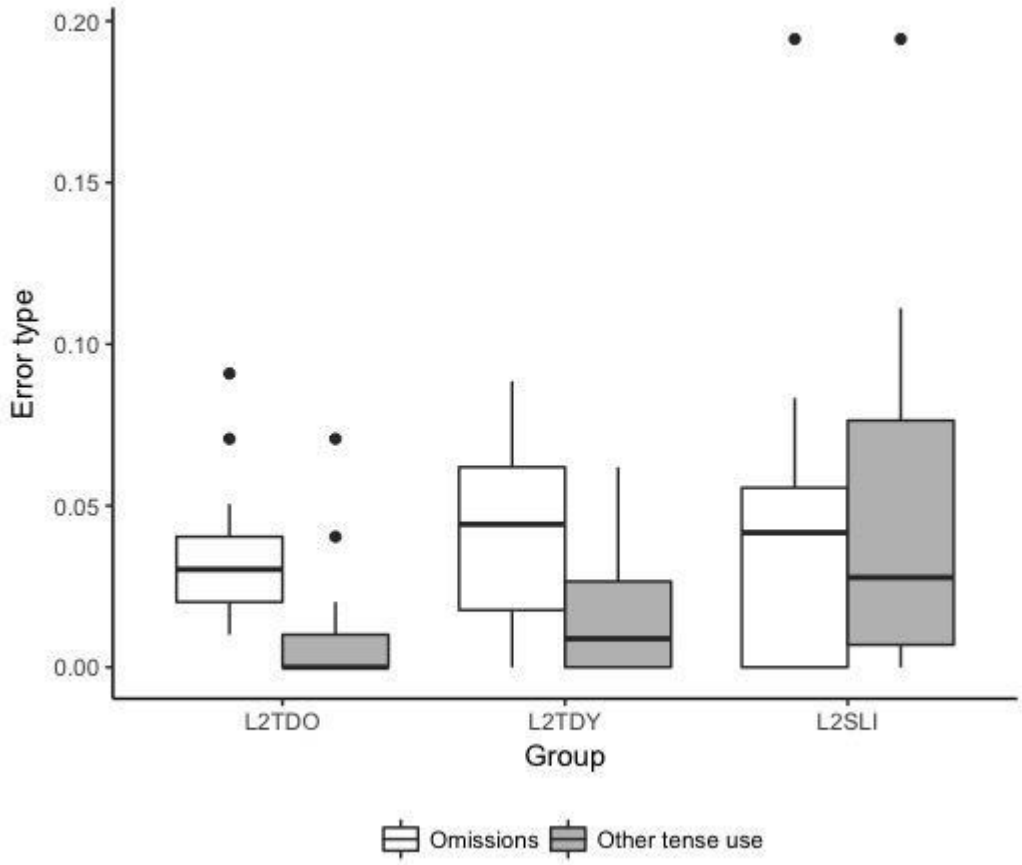


Figure 3. Accuracy of regular past tense for the L2-TLD_O, the L2-TLD_Y and the L2-SLI_Y children.

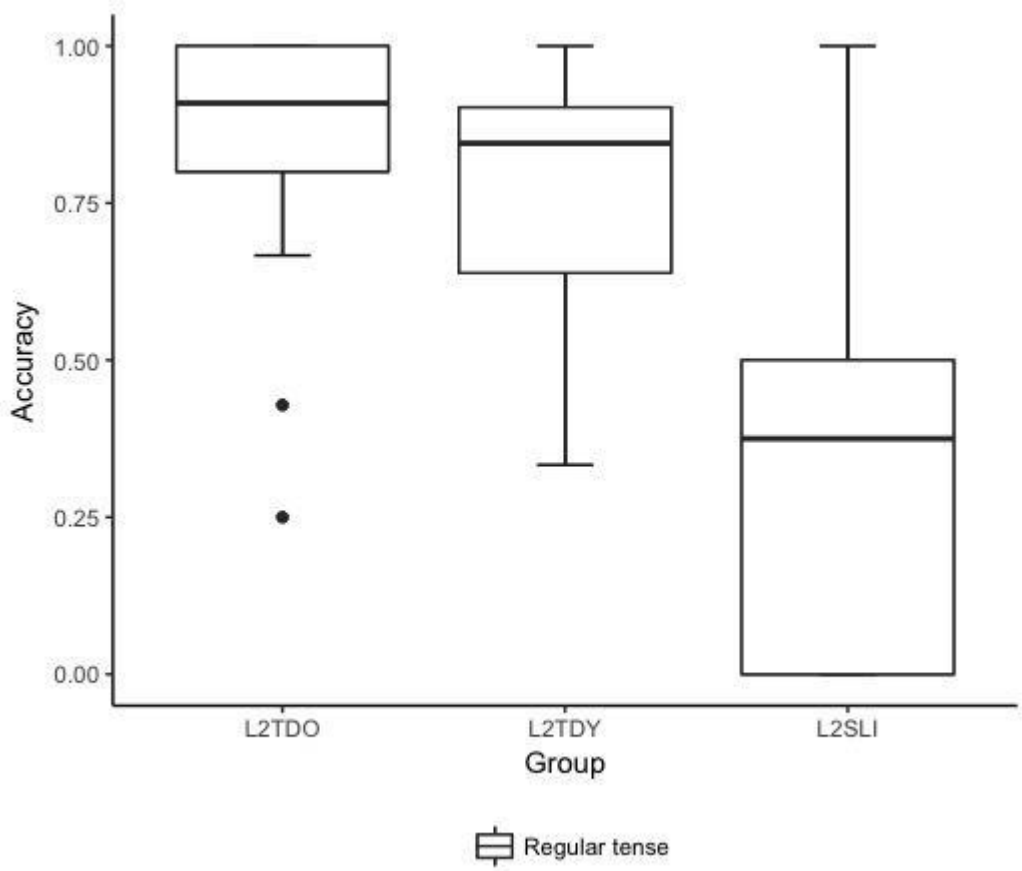


Figure 4. Error types of regular past tense for the L2-TLD_O, the L2-TLD_Y and the L2-SLI_Y children.

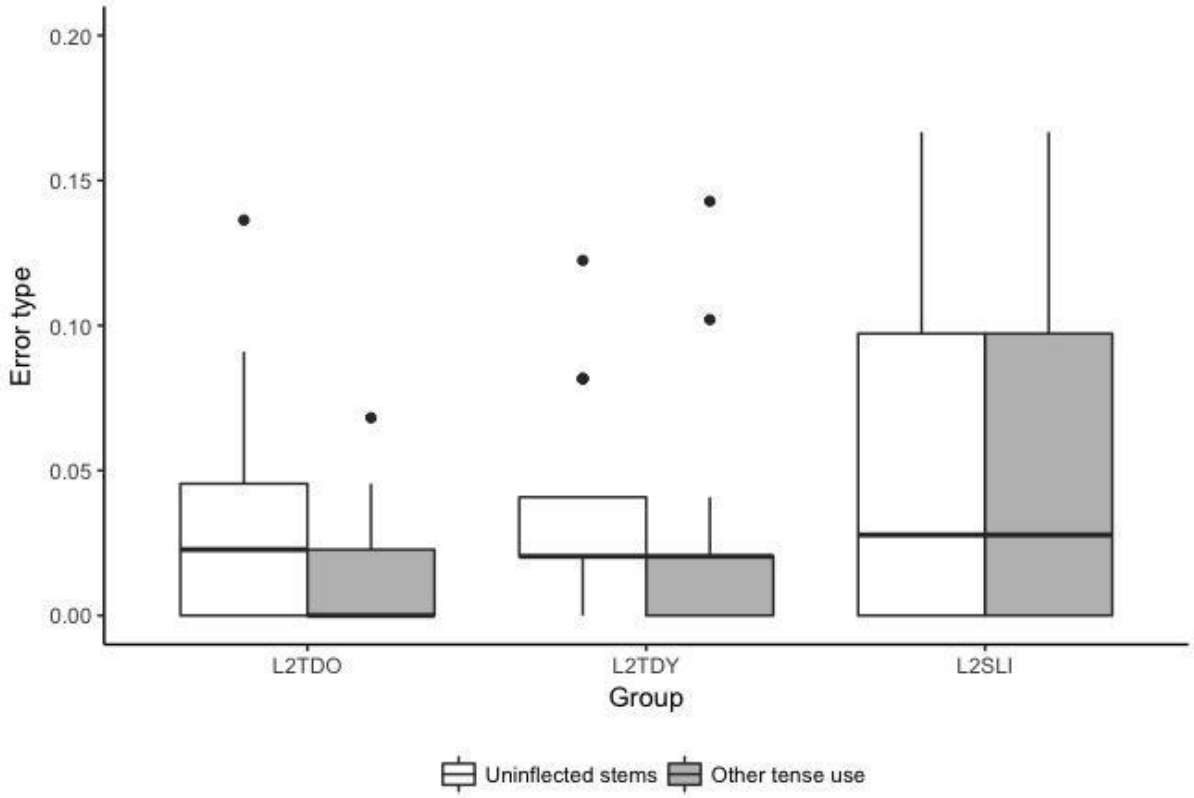


Figure 5. Accuracy of irregular past tense for the L2-TLD_O, the L2-TLD_Y and the L2-SLI_Y children.

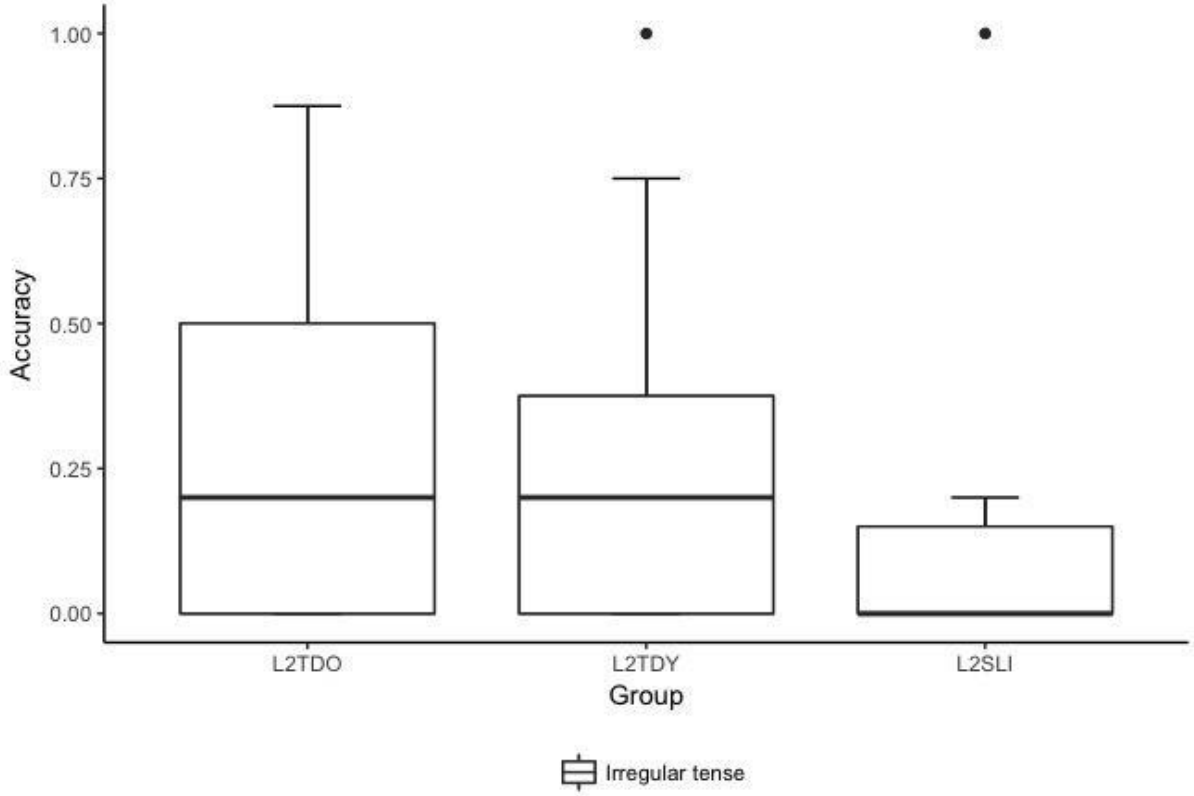
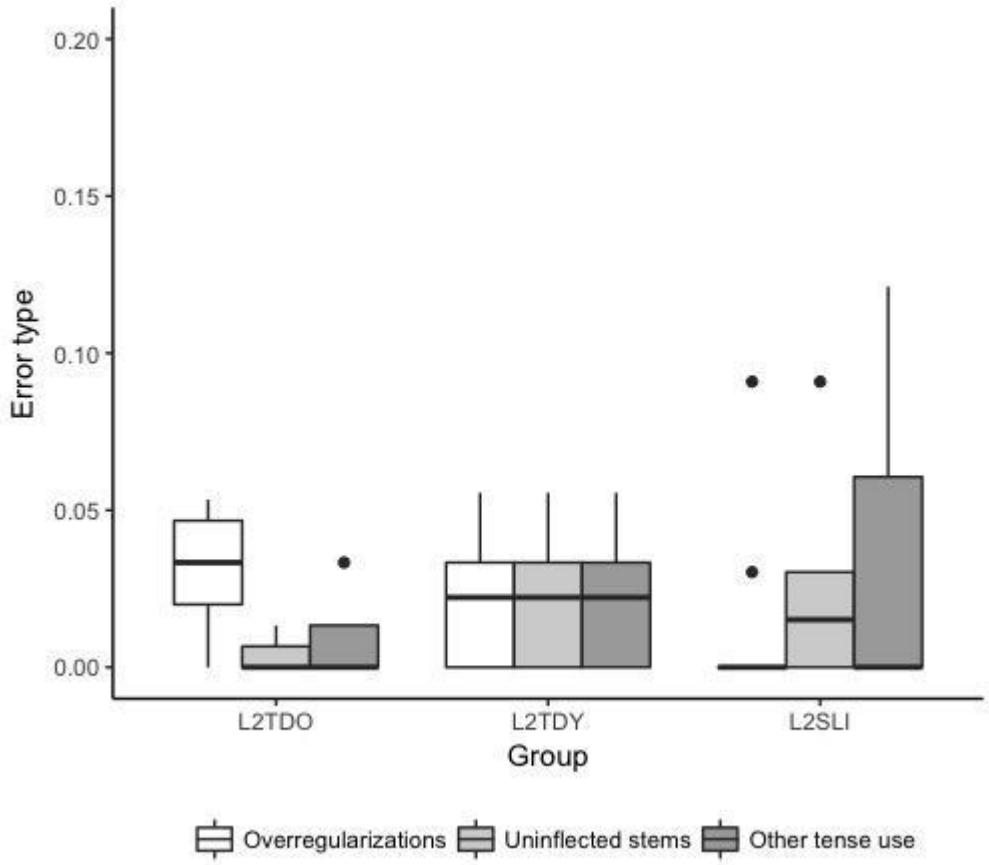


Figure 6. Error types of irregular past tense for the L2-TLD_O, the L2-TLD_Y and the L2-SLI_Y children.



Appendix. Verbs in the TEGI probes (Third person Singular, Past tense) and verbs used by the English L2 children.

Probe type	TEGI stimuli (Rice & Wexler, 2001)	All verbs used by the children
3SG *–s*	dentist, police officer, firefighter, pilot, painter, baseball player, nurse, astronaut, dad, dancer	arrest, bat, blast, blow, brush, catch, check, clean, correct, dance, do, drive, eat, fix, fly, get, go, help, hit, keep, look, make, paint, play, protect, put, rescue, ride, say, see, spray, stop, squirt, take, tell, throw, walk, wash, watch
Regular	brush, clean, climb, jump, kick, lift, paint, pick, plant, tie	brush, carry, clean, climb, close, jump, kick, land, lift, like, look, paint, pick, plant, pull, reach, smell, splash, stop, tidy, tie, turn
Irregular	blow, catch, dig, eat, give, make, ride, write	blow, build, catch, dig, eat, fall, feel, fly, get, give, go, grow, hold, lose, make, ride, smell, stand, think, throw, write

Note. TEGI=Test of Early Grammatical Impairment, L2=second language.