LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

Effects of Input Properties, Vocabulary Size, and L1 on the Development of Third Person Singular –*s* in Child L2 English

Elma Blom
University of Amsterdam

Johanne Paradis
University of Alberta

Tamara Sorenson Duncan
University of Alberta

This study was designed to investigate the development of third-person singular (3SG) –*s* in children who learn English as a second language (L2). Adopting the usage-based perspective on the learning of inflection, we analyzed spontaneous speech samples collected from 15 English L2 children who were followed over a 2-year period. Assessing the contribution of a wide range of predictors, we show that word frequency, allomorph, lexicon size, inflectional properties of the first language (L1), and months of exposure to English all have impact on English L2 children's use of 3SG –*s* in obligatory contexts. This study enhances both our understanding of the development of 3SG –*s* and of child L2 acquisition. The outcomes support a usage-based approach to learning inflection and emphasize the importance of a multifactorial analysis of language development.

Keywords child L2 acquisition; usage-based theory; inflection; frequency; lexicon-grammar dependency; language transfer; single-route models

This research was supported by grants from the Niels Stensen Foundation and a Marie Curie International Fellowship within the 7th European Community Framework Programme (IOF-219276) awarded to Elma Blom and grants from the Social Sciences and Humanities Research Council of Canada (SRG# 410–2002-0093) and the Alberta Heritage Foundation for Medical Research (Establishment Grant #20010034) awarded to Johanne Paradis.

Correspondence concerning this article should be addressed to Elma Blom, Universiteit van Amsterdam, Spuistraat 134, 1012 VB, Netherlands. Internet: w.b.t.blom@uva.nl

DOI: 10.1111/j.1467-9922.2012.00715.x

Introduction

Since the studies of Dulay and Burt (1973, 1974) it has been known that children learning English as a second language (L2) make errors with grammatical morphemes and that one of the last morphemes to be acquired is the third person singular (3SG) –*s* inflection. This inflection marks the third person of present habitual verbs, for example, "the boy walks to school every day" vs. "the boy is walking to school [right now]" or "I walk to school every day." More recent research that brought English L2 children's difficulties with 3SG –*s* to the focus of attention are Haznedar (2001), Ionin and Wexler (2002), and Paradis, Rice, Crago, and Marquis (2008). While prior research has documented learners' difficulty in acquiring 3SG –*s*, few studies have aimed at explaining what factors impact its development, in child L2 learners in particular. Using linear mixed effects modeling we assess in this study the contribution of a range of factors. In order to single out potentially relevant factors, we adopt the usage-based perspective on learning inflection as developed in the work of Bybee (1995, 2001, 2008, 2010) and, in so doing, test predictions derived from this model as part of this study.

A Usage-Based Approach to Inflection Learning

In this study, the development of 3SG –*s* inflection in child L2 English is investigated in terms of Bybee's usage-based Network Model (1995, 2001, 2008, 2010). We will henceforth refer to this theory with NM. The NM typically takes a domain-general perspective and hence assumes that learners apply the same mechanisms for learning inflection as for learning other cognitive skills, for example, "categorization, chunking, rich memory storage, analogy and cross-modal association" (Bybee, 2010, p. 7). Crucially, instances of language use impact the cognitive representation of linguistic information. The assumptions of the NM contrast with a domain-specific perspective, according to which inflection is acquired through processing mechanisms specific to language and, particularly, specific to grammar (cf. Pinker & Ullman, 2002). Below, we will explain the NM in greater detail, and how it applies to learning inflection.

According to the NM, the acquisition of inflection is dependent on lexical strength. Lexical strength can be viewed as a measure of the processing burden of any given word. Each time a word is encountered it leaves a small trace in the lexicon. A word that is encountered more often will become more ingrained in the lexicon, and its lexical strength will increase. Each subsequent encounter of that word will require less processing effort on behalf of the processor, and as the word's lexical strength increases, it will be more easily retrieved. In the NM, different inflectional forms of a verb are supposed to be stored in their own right as multimorphemic units. Thus, "talk," "talks," "talked" and "talking" would all be separate phonological/semantic entries. Therefore, multimorphemic storage combined with the processor's sensitivity to frequency means that word forms with higher token frequencies are likely to be used correctly by language learners, whereas word forms with low token frequencies are prone to errors. But lexical strength is also related to type frequency. Whereas token counts tally the raw number of occurrences for a particular word form, type frequency takes into account the number of roots with which a certain inflection is used (e.g., "talks," "likes," "brings," "plays," and so on). Type frequency determines the strength of a schema, which is an abstraction over various, specific lexical instantiations with semantic and/or phonological similarities. A suffix with a high type frequency has a strong schema and will be used with novel roots. A frequent suffix that appears with a limited set of roots has a high token frequency but a low type frequency. Such a suffix will not be used productively, and its emerging correct use by learners is dependent on token frequency alone. Apart from type frequency, the degree of schematicity, determined by the similarity among instantiations, is considered a determinant of productivity.

Frequency Effects

The NM predicts that the development of 3SG –*s* inflection is determined by how often a verb occurs with this inflection in the input, because each instance of use will strengthen a learner's memory representation (Bybee, 2008, p. 219). Children are thus expected to be more accurate with verbs that occur frequently with 3SG –*s* in the input. In this study, we will refer to this factor as "word frequency" (i.e., token frequency), which we will distinguish from "lemma frequency." Lemma frequency denotes the frequency of a verb in the input regardless of its inflected form. Thus, the lemma frequency of the verb "talk" is the sum of the word frequencies of "talk," "talks," "talked," and "talking." Lemma frequency is not expected to facilitate the development of 3SG –*s*, and might even cause more errors for 3SG –*s*, in particular if word frequencies other than 3SG –*s* contribute disproportionally to lemma frequency. For instance, a verb with a high word frequency for 3SG –*s* but an even higher word frequency for past tense –*ed* has a stronger connection with past tense –*ed* than with 3SG –*s*, which may inhibit retrieval of [verb+3SG –*s*].

Currently, it is unknown to what extent input distributions affect L2 children's performance with 3SG –*s*. In this study we will fill this empirical gap by examining the development of 3SG *–s* in child L2 English in relation to characteristics of the language they are exposed to. Studies that addressed this issue in first language (L1) acquisition report inconsistent outcomes. In an experimental setting, significant correlations emerged between the form in which children had learned novel verbs (with or without 3SG *–s*) and how they produced these verbs after the training session (Finneran & Leonard, 2010; Theakston, Lieven, & Tomasello, 2003). It is, however, questionable if these experimental findings can be generalized to language acquisition in real life, because naturalistic data have not shown any correlations between L1 children's accuracy with 3SG *–s* and frequency distributions in the input (Song, Sundara, & Demuth, 2009). However, as Song et al. (2009) point out, it is possible that the frequencies in the corpus they used for input estimation are not sufficiently representative. Therefore, in the present study, we analyzed two corpora of spoken language; one very large corpus which is possibly less representative of the input to the L2 children in our study, and one smaller, but highly representative corpus.

Variation Between Allomorphs

According to the NM, learners organize associations between words in schemas based on semantic and phonological similarities (Bybee, 2001, pp. 21–22). Such schemas capture semantic and phonological generalizations over words, morphemes, syllables, sounds. Schemas have been argued to explain developmental variation between the allomorphs of the English plural suffix *–s* and past tense *–ed*, in particular the high omission rates of the plural ending /ɪz/ and the past tense marker /ɪd/ in L1 English preschool children (Berko, 1958; Köpcke, 1998; Bybee, 2007). The relatively frequent drop of /ɪz/ and /ɪd/ suggests that children work with schemas which state that nouns ending in /s/ or /z/ are plural and verbs that end in /t/ or /d/ are acceptable past tense forms. These schemas would also apply to singular nouns ending in /s/ or /z/ (e.g., "box") and to present tense verbs ending in /t/ or /d/ (e.g., "rate"). If children are less concerned with the addition of a suffix than with how the general shape of the word fits the pattern, /ɪz/ and /ɪd/ are likely to be omitted with such nouns and verbs (Bybee, 2007). This effect will persist if these nouns/verbs in plural/past tense form do not have high word frequencies and if /ɪz/ and /ɪd/ have weak schemas due to low type frequencies.

Previous studies have investigated variation between the allomorphs of plural *–s* and past tense *–ed* (Berko, 1958; Köpcke, 1998; Bybee, 2007), but no study has looked at developmental variation between 3SG /z/, /s/, and /ɪz/. Therefore, in this study, we will examine whether or not different developmental patterns emerge for the allomorphs of 3SG –*s* in child L2 English, and if these patterns follow the predictions of a usage-based model of learning inflection.

Lexicon and Grammar

In previous child L2 studies, it has been argued that observed associations between lexicon size and grammatical skills support domain-general accounts that assume that one unitary processing mechanism underlies the lexicon and grammar (Simon-Cereijido & Gutiérrez-Clellen, 2009; Kohnert, Pui Fong, & Conboy, 2010). The NM, sharing this assumption, may make the same prediction. In the NM, an inflection with a high type frequency will have a strong schema and will be more productive than an inflection with low or medium type frequency. The relation between type frequency and productivity is supported by Marchman and Bates's (1994) observation that the onset of overregularization of English regular past tense inflection is related to an increase in children's lexicon size (Bybee, 2001, p. 120). Along the same lines, we may expect that lexicon size, and in particular verb lexicon size, may affect children's schemas related to 3SG –*s*. Consequently, lexicon size can determine how well L2 children can use this inflection in obligatory contexts across different verbs, which, in turn, can affect their overall accuracy on 3SG –*s*.

Most studies examined the relation between the lexicon and grammar using general measures based on standardized tests (for L1, see Bates et al., 1994; Caselli, Casadio, & Bates, 1999; Fenson, Dale, Reznick, Bates, Thal, & Pethic, 1994; for simultaneous bilinguals, see Conboy & Thal, 2006; Marchman, Martínez-Sussman, & Dale, 2004; for sequential bilinguals, see Cobo-Lewis, Eilers, Pearson, & Umbel, 2002; Kohnert et al., 2010). A few studies compared more specific lexical and grammatical properties, such as verb lexicon size and past tense in L1 English (Marchman & Bates, 1994) and verb lexicon size and verb argument structure in Spanish-English successive bilingual school-age children (Simon-Cereijido & Gutiérrez-Clellen, 2009).

To date, no study has investigated whether the association between the lexicon and grammar found in other studies extends to 3SG –*s*. Moreover, available studies correlated group data and did not compare lexical and grammatical measures within individual children, which would be a more appropriate method (Rispoli, Hadley, & Holt, 2009). In the present study, we investigated the impact of a child's lexicon size, including more general measures of a child's lexicon size as well as verb lexicon size in production, which is a more specific measure. The use of mixed linear regression allowed us to look at the effect of (verb) lexicon size on the child's accuracy with 3SG –*s* within children.

L1 Transfer

The NM applies to morphological processing, diachronic changes, and to language acquisition. Bybee (2008) discusses the NM in relation to L2 acquisition. However, a detailed account of L1 transfer in this model is currently lacking, which may reflect the fact that application of a usage-based perspective to issues in L2 research and learning is still in its infancy (Tyler, 2010, p. 283).

In a brief section on transfer, Bybee (2008, p. 232) states that L2 learners can transfer constructions from their L1 to their L2 on the basis of similarity. Thus, although according to the NM schemas emerge from words, it may be possible that abstract features such as person, number, and tense are at some level disconnected from these lexical items and transferred to the lexicon of an L2. The possibility to transfer more abstract language representations seems consistent with other usage-based models. Gathercole (2007) concludes that bilingual children may transfer patterns across languages if these patterns are sufficiently similar to invite transfer and sufficiently abstract so that transfer is not dependent on lexical-specific information. MacWhinney (2008a) posits that "although arbitrary forms and classes cannot transfer between languages, the grammatical functions underlying affixes can" (p. 353). From this it may be inferred that English L2 children's acquisition of 3SG –*s* may be supported by L1 schemas for 3SG inflection, whereas English L2 children who do not have such schemas from their L1 lack prior knowledge that can help them identifying 3SG –*s*.

Transfer in a usage-based framework may also have to do with attention and perception, that is, with cognitive mechanisms for processing language that might be honed to L1 properties (Ellis, 2006, 2008). The processing and perception of verbs in the L2 will be shaped by L1 schematic representations of verbs. In the case of isolating L1s, verbs would lack associations with features like person and number. As a result, children with isolating L1s might not initially attend to, or perceive, 3SG –*s* in English. Children whose L1s have verb inflection that encodes person and number features, in contrast, will have strong associations between verb schemas and person and number features that could carry over to their acquisition of verbs in the L2. These children may attend to the end of verbs because in their L1 this position provides information for interpreting the subject of the sentence and to assign a temporal interpretation to the proposition expressed.

Less than a handful of studies have looked at the child L2 development of 3SG –*s* as a function of the L1. McDonald (2000) reported long-lasting effects of transfer in the grammaticality judgment data of Vietnamese and Spanish L1 learners whose exposure to English started during childhood and after puberty. Inflectional morphology, including 3SG –*s*, was one domain where the Vietnamese L1 learners performed more poorly than did the Spanish L1 learners, even when they were exposed in childhood to their L2 (English). This observation suggests that having agreement inflection in the L1 determines performance with 3SG –*s*, given that Spanish has a rich system of agreement inflection whereas Vietnamese has no agreement inflection. Studies of the production of English verb inflection report mixed outcomes. Paradis (2011) observed that children with an isolating L1 (Mandarin/Cantonese) were less accurate on –*ed* and –*s* than children with inflecting L1 (Hindi/Punjabi/Urdu, Spanish, Arabic). Dulay and Burt (1974) found that children whose L1 was a Chinese language were, as a group, less accurate with 3SG –*s* than children whose L1 was Spanish, but these researchers did not collect data on how long the children were exposed to English, and it is possible that the Chinese L1 children had different levels of exposure to their L2. Paradis (2005) and Paradis et al. (2008) did not find differences based on the inflectional properties of the L1 in children's accuracy with 3SG –*s*; however, the children were studied after just 1 year of exposure to L2 English, and it is possible that L1 effects emerge over time. Therefore, further research on the influence of L1 typology on child English L2 acquisition of inflection is needed.

Research Questions

The goal of this study was to examine which factors determine the development of 3SG –*s* in child L2 English. In our review of the literature, we have singled out a number of factors which are, according to the NM, potentially relevant for explaining patterns in the development of 3SG –*s*. In order to investigate the contribution of these factors, we formulated the following four research questions for our study:

(1) Do frequency distributions in the input impact on the acquisition of 3SG –*s* in child L2 English?
(2) To what extent do L2 English children vary between the allomorphs of 3SG –*s*, and what conditions this variability?
(3) Is the development of 3SG –*s* in child L2 English predicted by the children's vocabulary size?
(4) Do inflectional properties of the L1 influence the acquisition of 3SG –*s* in child L2 English?

The roles of frequency distributions, allomorph, vocabulary size, and L1, if addressed at all in relation to the development of 3SG –*s*, have yielded conflicting outcomes in previous research. Hence, by investigating these factors, our study enhances understanding of how child L2 grammar unfolds and if it follows the predictions of a usage-based approach to learning inflection.

Method

Participants

We analyzed spontaneous language sample data from 15 English L2 children living in an English majority city: Edmonton, Canada. The sample consisted of 11 boys and 4 girls. The children spoke one of the following languages as a L1: Cantonese ($N = 2$), Mandarin ($N = 6$), Cantonese/Mandarin ($N = 1$), Romanian ($N = 1$), or Spanish ($N = 5$). Twelve of the children were recently arrived immigrants and three were born in Canada. The children who were born in Canada received almost exclusive exposure to the home language (L1) and did not begin receiving consistent exposure to English until they started preschool or school. Children were recruited through agencies that offer assistance to newcomer families and through government-sponsored English-language classes for adults. This study has a longitudinal design, thus five spontaneous language samples were collected from the children at 6-month intervals over 2 years. For English L2 children, data existed for each round of data collection. Two children were only tested at rounds 1 and 2, and one child was only tested at rounds 3, 4, and 5; in total, 67 language samples were available. The children in this study participated in previous studies, including Paradis (2005), Paradis et al. (2008), Golberg, Paradis, and Crago (2008), and Zdorenko and Paradis (2008). The other studies looked at these children as part of a larger group, and 3SG –*s* was examined only at round 1 in Paradis (2005) and Paradis et al. (2008), and for different research questions.

Table 1 gives a descriptive overview of the children's ages in months and months of exposure to English (MOE) at the five rounds. More participant information is provided after we have described the procedures; also the measure of exposure to English is described in the Procedures section.

Procedures

Data on the children's performance with 3SG –*s* come from analyses of spontaneous language data collected during freeplay sessions of 30–45 minutes each. These language samples consist of a conversation between the child

Table 1 Age and exposure to English (MOE) in months across rounds

		Age			MOE		
Round	*N* children	*Min-max*	*Mean*	*SD*	*Min-max*	*Mean*	*SD*
1	14	45–81	67	7.60	4–18	9	3.75
2	14	61–86	73	8.03	10–24	16	3.86
3	13	67–93	79	7.71	16–30	22	3.92
4	13	72–99	86	8.06	22–35	28	3.85
5	13	79–105	92	8.40	28–42	35	4.54

Note. *SD* = Standard deviation; *Min-max* = Minimum-maximum.

and an English-speaking research assistant. Research assistants allowed each child to guide the conversation to ensure that the child was as engaged in the conversation as possible. Each research assistant, however, was also provided with a list of questions that could be used to prompt the child to talk if there was a lull in the conversation.

The language samples were transcribed in CHAT format to allow analysis using CLAN programs (MacWhinney, 2000), and coded for the use and accuracy of 14 grammatical morphemes in obligatory context (Brown, 1973), including 3SG –*s*, by native-speakers of Western Canadian English; only 3SG –*s* results will be reported in this article. Twenty percent of the corpus was re-transcribed and coded by a second research assistant. These transcripts were then compared to the original transcripts and reliability scores were generated for both the transcription and coding. For the transcription, the inter-rater reliability (measured by words) was 94.5% (*SD* = 2.6%; minimum–maximum = 89.5%– 98.1%). For the coding, the inter-rater reliability was 91.1% (*SD* = 3.0%; minimum–maximum = 85.2%–96.6%). All disagreements were settled by consensus between the two transcribers.

For our first research question, characteristics of frequency distributions in the input are relevant. In order to estimate the impact of word frequency and lemma frequency on children's performance with 3SG –*s*, we consulted two corpora. The Corpus of Contemporary American English (COCA) is a very large, tagged corpus of North American English. The spoken word subcorpus that we searched consisted of 83 million words at the time we consulted it. The COCA corpus is not necessarily highly representative of the variety the children in our sample are exposed to. For this reason, a small corpus, henceforth the Edmonton English Language Learners (ELL) corpus, was compiled based on video-taped spontaneous language samples gathered during semi-structured freeplay sessions between research assistants and L2 children, as part of an ongoing cross-sectional study with children from immigrant families living in Edmonton (being conducted by the second author). The Edmonton ELL corpus consists of both the speech of bilingual children in Western Canada and of the research assistants addressing the children, who were native speakers of Western Canadian English. While it is significantly smaller (400,824 words) than the COCA corpus, it is more representative of the language that the children in our sample are exposed to, because it is Western Canadian English and, more importantly, because a substantial part of the speech in the Edmonton ELL corpus (64% of the utterances) is speech from research assistants addressed to children very similar to the children studied here. The other part of the corpus (36%) is speech from ELL children, which is representative of speech that the present participants will be exposed to via their (ELL) peers.

All utterances in the Edmonton ELL corpus used by children and assistants were transcribed in CHAT format, and further analysed using CLAN commands. The files were first morphologically tagged (with MOR), and subsequently disambiguated with POST. The POST command creates tags for English that are correct in 95% of the cases (MacWhinney, 2008b). Using FREQ, we created frequency lists for all verbs occurring in the Edmonton ELL corpus, with the purpose of calculating word frequency and lemma frequency in the input.

As mentioned in our review of the literature, word frequency refers to how often a verb appears with 3SG inflection in the input. Lemma frequency refers to how often this verb appears in the input, irrespective of its inflectional form. Lemma frequency included bare stems ("walk"), 3SG –*s* ("walks"), past tense ("walked"), past participles ("walked"), and progressive participles ("walking"). We counted, for each verb used by the children in 3SG contexts, its word frequency and lemma frequency in both the COCA corpus and in the Edmonton ELL corpus, and entered this number in the dataframe. Word frequencies and lemma frequencies in the Edmonton ELL corpus ranged from 0 to 552 and from 0 to 2,919, respectively. Word frequencies and lemma frequencies in the COCA corpus ranged between 0 – 21,073 and between 0 – 327,901, respectively. Because the Edmonton ELL corpus contained speech from ELL children, this corpus contained errors, for example, omissions of 3SG –*s*. These errors were included in lemma frequency. To a certain extent the children in the study will be exposed to such errors, but in order to ensure that including the child data did not bias the corpus, we calculated correlations between the whole corpus and a subpart of the corpus with only the speech of the research assistants. Both word frequency ($r(196) = 0.99$, $p < .0001$)

and lemma frequency ($r(196) = 0.93$, $p < .0001$) correlated very strongly, indicating that the frequencies counts based on the whole corpus are not biased.

In addition to word and lemma frequency, each verb was coded for allomorph type, distinguishing between verb stems that end with a sibilant (/ɪz/), verb stems with a final voiceless consonant (/s/), and verbs that end with a final voiced consonant or a vowel (/z/). Variation among allomorphs pertains to the second research question addressed in this study.

The language samples were analyzed for children's use of 3SG –*s*, but the samples also provided information on the children's verb lexicon size, which is relevant for the third research question. Verb lexicon size was calculated by taking the number of different verbs per 100 utterances (we call this measure Number of Different Verbs [NDV]). NDV was calculated using the FREQ command in CLAN; calculations were based on the last 100 utterances of each transcript, because children generally were more comfortable and talkative later in the session. The language samples provided further information on the children's expressive vocabulary, as indexed by the lexical diversity measure *D* (Malvern & Richards, 2002), and their expressive grammar, estimated by mean length of utterance in words (MLUW). The lexical diversity measure *D* has been developed to overcome effects of sample size and is based on the probability of new vocabulary being introduced in longer and longer samples of speech or writing. *D* and MLUW were calculated using the CLAN commands VOCD and MLU (MacWhinney, 2000). Each child was also given the Peabody Picture Vocabulary Test (PPVT-III, Dunn & Dunn, 1997), which is a measure of children's receptive vocabulary size. The PPVT consists of a series of picture arrays shown to the child and the child is asked to point to the picture that matches the word spoken by the experimenter. NDV, *D*, and the raw PPVT scores were included as independent variables in this study to examine whether or not children's performance on 3SG –*s* was predicted by size of the receptive lexicon.

In order to answer the fourth research question, the children were assigned to groups based on inflectional properties of the L1. Spanish and Romanian L1 children were assigned to the group of children with richly inflecting L1s because Spanish and Romanian both inflect for person, number, and tense (Butt & Benjamin, 2004; Gönczöl-Davies, 2007). Mandarin and Cantonese L1 children were assigned to the group of children with isolating L1s because both Mandarin and Cantonese are isolating languages that do not have tense and agreement inflection (Matthews & Yip, 1991; also Lin, 2001). Table 2 lists the receptive (PPVT) and expressive vocabulary (*D*) scores as well as verb lexicon size (NDV) and the expressive grammar scores (MLUW) of the two L1 groups.

Table 2 Participant characteristics[a]

		Inflecting L1 ($N = 6$)	Isolating L1 ($N = 9$)
Number of different verbs (NDV)	*M* (*SD*)	27.35 (9.35)	23.79 (5.84)
	Min-max	12–46	12–40
Receptive vocabulary (PPVT)	*M* (*SD*)	89.32 (28.29)	84.53 (26.11)
	Min-max	19–144	29–135
Expressive vocabulary (*D*)	*M* (*SD*)	68.92 (16.68)	66.03 (17.38)
	Min-max	27.50–101.44	36.10–99.13
Expressive grammar (MLUW)	*M* (*SD*)	3.93 (0.59)	3.64 (0.77)
	Min-max	2.57–5.02	2.09–5.52
Age	*M* (*SD*)	83.26 (10.34)	79.44 (12.77)
	Min-max	61–101	54–105
Months of English (MOE)	*M* (*SD*)	22.4 (8.90)	24.5 (10.20)
	Min-max	5–40	4–42
Mother's education (MOTED)	*M* (*SD*)	14.6 (2.50)	16.4 (0.80)
	Min-max	9–17	15–18
English in the home	*M* (*SD*)	0.43 (0.06)	0.43 (0.07)
	Min-max	0.38–0.58	0.30–0.50

Note. *M* = Mean; *SD* = Standard deviation; *Min-max* = Minimum-maximum. [a]The data are collapsed across time.

The research assistant, with the assistance of an interpreter/cultural broker, collected background information on individual children through an oral interview with parents. The questionnaire provided, among other things, information about children's age, months of exposure to English, mother's level of education, and amount of English spoken at home. Table 2 also summarizes this information for the two L1 groups. Mother's level of education was calculated using years of formal education: Six years is primary school, 12 years corresponds to the end of secondary school, and years greater than 12 refers to postsecondary education. Information about languages spoken at home and how often these languages were spoken was documented using a 5-point scale (1 = *native language only*, 5 = *English only*). The father's, mother's, and sibling's language use with the child, and the child's language use with these family members, were recorded on separate 5-point scales. The values from each scale were added together and divided by the maximum number of points from the total number of scales to calculate a proportional value for the amount of English spoken in the home.

Children with an inflecting and isolating L1s had similar verb lexicon sizes ($W = 555, p = .614$; based on Wilcoxon rank-sum test), receptive vocabularies

Table 3 Data analyzed across rounds

Round	*N* children	*k* contexts	Contexts across children *Min-max*	*Mean*	*SD*	*k* correct 3SG –*s*	*k* omitted 3SG –*s*
1	14	168	6–18	10.80	3.70	28	140
2	14	177	7–18	11.90	3.70	88	89
3	13	220	3–35	16.90	9.30	145	75
4	13	266	7–43	20.50	11.80	222	44
5	13	294	7–39	24.70	10.70	248	46

Note. k = number of cases for analysis; *SD* = Standard deviation; *Min-max* = Minimum-maximum.

($W = 504.5, p = .8856$), expressive vocabularies ($W = 576.5, p = .4327$), and expressive grammar skills ($W = 657, p = .0659$), when the data for all children across the five rounds are taken into account. Furthermore, the children in the two L1 groups were of identical ages ($W = 553, p = .633$), had similar length of exposure ($W = 419$, $p = .2065$), their mothers did not differ in level of education ($W = 36.5, p = .248$), and the same amount of English was spoken in the home ($W = 27.5, p = 1.0$). Thus, comparisons between the L1 groups for the use of 3SG –*s* in English are unlikely to be confounded by between-groups differences based on other sample characteristics.

Analyses

The outcome variable of this study is the use of regular 3SG –*s* inflection in obligatory contexts. For each obligatory context in the transcribed speech samples, it was coded whether 3SG –*s* was realized or omitted. Contexts other than 3SG were not taken into account; hence, children's incorrect use of 3SG –*s* in other contexts (e.g., "I sees two leg"), was not evaluated ($k = 52$). Occasionally, a child used a past tense in contexts in which 3SG –*s* was obligatory; such responses were not taken into account ($k = 39$). The verbs "be," "have," "do," and "say" were excluded ($k = 385$), because for these verbs 3SG is irregular (i.e., suppletion or stem alternation). Table 3 shows the data included in the analyses for each round and child. In total, we analyzed 731 correct uses of 3SG –*s* and 394 omissions. Each child produced at least six obligatory contexts of 3SG –*s* at each round, except for one child for whom in the third round only three contexts were available.

In order to investigate the effect of input frequencies, allomorphs, lexicon size, and L1 on L2 children's accuracy with 3SG –*s*, we analyzed the data statistically using a generalized linear mixed model with a logistic link function for correct use and omission of 3SG –*s* in obligatory contexts, and binomial variance. Logistic regression is particularly suited for analyzing the contribution of multiple factors on a discrete outcome variable, such as use or omission of 3SG –*s*. Regression on proportions has various mathematical drawbacks. Therefore, in logistic regression the log odds ratio is used as the dependent variable, instead of proportions. In our study, the odds ratio is the ratio of the probability of realized 3SG –*s* inflection over the probability of omitted 3SG –*s* inflection in obligatory contexts; the log odds ratio is the log transformed odds ratio.

Advantages of logistic regression include its robustness, which is important in the case of unequal observations, missing data, and small samples. An additional advantage of mixed regression modeling is that random-effect variables are included, next to fixed-effect variables. Random-effect variables, in contrast to fixed-effect variables, are assumed to be measured with error, their values come from a larger population and are intended to generalize to this larger population, and they will have different levels in each new study. For instance, the effect of the participating children is typically considered a random-effect variable. This is because the sample of children is drawn from a larger population, and each participating child has unknown properties that will influence the measurements. Hence, the effect of participating children cannot be measured without error and if a new study is undertaken, other participants will be included, with again unknown properties. The same characteristics apply to the sample of verbs included in this study: As with children, the effect of the verbs included in the study cannot be measured without error, because each verb has properties that are unknown.

Therefore, in this study, both the factors Child and Verb are modeled as random-effect factors, which have a mean of zero and unknown variance. By including the random-effect variables Child and Verb, we ensured that the predictions of the optimal regression model can be generalized to the population from which the samples are drawn; hence, the model's predictions will likely extend to children and verbs other than those analyzed in this study. For further explanations of the advantages of mixed logistic regression over analyses of variance or ordinary logistic regression, based on psycholinguistic data and linguistic corpora, we refer readers to Baayen (2008) and Jaeger (2008).

Fixed-effect predictors in our study that relate to the first three research questions are NDV, *D*, PPVT, Word frequency (COCA corpus, Edmonton ELL corpus), Lemma frequency (COCA corpus, Edmonton ELL corpus), and Allomorph. The frequency data were transformed into logarithmic scale backing off from zero to prevent division by zero. The fixed effect predictor L1 distinguished between children whose L1 had both person and number inflection and children whose L1 had no inflection. In addition, we included MOE as a measure of L2 development.

Word frequency and lemma frequency in the COCA corpus showed a correlation of $r(196) = 0.66$ ($p < .0001$), whereas for the Edmonton ELL corpus a correlation of $r(196) = 0.36$ ($p < .0001$) was obtained. Also, the correlation between NDV and PPVT was significant ($r(67) = 0.32, p < .01$), as were the correlations between NDV and D ($r(67) = 0.39, p < .01$) and between PPVT and D ($r(67) = 0.44, p < .001$). Finally, significant correlations existed between MOE and NDV ($r(67) = 0.31, p < .05$), MOE and PPVT ($r(67) = 0.57, p < .0001$) and MOE and D ($r(67) = 0.33, p < .01$).

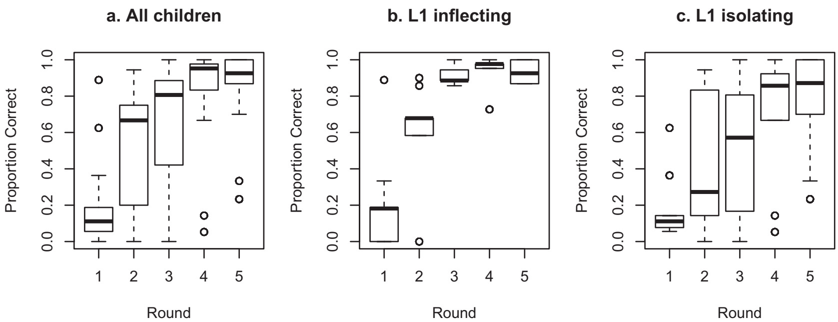
The strength of a correlation might make it difficult, if not impossible, to tease the effect of two predictors apart. Therefore, we decorrelated each pair of correlated factors by predicting one variable using the other as a predictor variable and entered the residuals for the predicted variable (i.e., the variation that could not be explained by the second predictor) as the predictor variable into the regression model. This method avoided overlap between two predictors.

Regressing children's performance on 3SG –*s*, we used the following procedures. First, we used the technique of backwards elimination. Starting with a full model that overfitted the data, we removed predictors that were not significant in a step-wise fashion. Second, from a simple model with only main effects arrived at from the backwards elimination, we gradually increased the model complexity by entering interactions between predictors, and we also squared terms for possible non-linear effects. Nested models that differed in complexity were compared using maximum likelihood ratio tests. The result of these procedures was a final model that was both as simple and as precise as possible. Because for logistic regression no simple satisfactory measure of goodness of fit exists, such as the R^2 for multiple regression models, the fit of the logistic model to the data was judged through the Concordance Index C (Chatterjee & Hadi, 2006). C ranges between 0.5 and 1; a C value above 0.8 indicates that the model performs well.

Results

Development Across Rounds

The box-plots in Figures 1a–c give an overview of the children's proportions of realized 3SG verb forms in obligatory contexts across the 5 rounds, with the lower and upper quartiles represented by the bottom and top of the box



Figures 1 a–c Mean proportions correct 3SG -*s* across rounds

and the median by the line in the middle of the box. A strong increase can be observed in use of 3SG –*s* at the beginning, with a flattening above the level of 75% correct at rounds 4 ($M = 0.83$, $SD = 0.27$) and 5 ($M = 0.84$, $SD = 0.24$). At round 4, the children have had on average 28 months of exposure to English ($SD = 3.85$; minimum–maximum = 22–35). On average, children with an inflecting L1 reached the level of 90% correct in round 3 ($M = 0.92$, $SD = 0.05$), maintaining this level in rounds 4 ($M = 0.94$, $SD = 0.09$) and 5 ($M = 0.93$, $SD = 0.06$). Children with isolating L1s did not reach a mean of 90% even by round 5 ($M = 0.77$, $SD = 0.29$), and furthermore, exhibited a greater degree of between-child variance at the final rounds when compared to the inflecting L1 group.

Mixed Logistic Regression

Stepwise elimination of predictors that did not reach significance resulted in a simple model with main effects for word and lemma frequency in the Edmonton ELL corpus, Allomorph, MOE, and NDV. Recall that word and lemma frequency in the Edmonton ELL corpus are correlated ($r(196) = 0.36$ ($p < .0001$), which means that it is difficult to separate the effects of the two predictors. Therefore, we reran the model with a predictor that contained the residuals of word frequency if the variation in word frequency is predicted by lemma frequency (see Analyses section). The correlation between the decorrelated word frequency predictor and the original word frequency predictor was significant ($r(196) = 0.52$, $p < .0001$), which indicates that the decorrelated predictor still predicted word frequency, albeit to the extent that it cannot be predicted from lemma frequency. In addition we increased the complexity of the model and, in so doing, obtained a model with main effects for word frequency in the Edmonton ELL corpus (using the residuals), NDV, and MOE, and interaction effects between L1 and lemma frequency in the Edmonton ELL corpus, and between L1 and Allomorph.

In the simple model with only main effects, the only measure of lexicon size that predicted children's performance on 3SG –*s* was NDV. However, by entering PPVT (receptive vocabulary) and *D* (lexical diversity in expressive vocabulary) in the more complex model with interactions, an effect of PPVT was obtained. Because PPVT was significantly correlated with NDV ($r(67) = 0.32, p < .01$), we predicted the variation in NDV using PPVT as a predictor; the correlation between NDV and NDVResid was significant ($r(67) = 0.90, p < .0001$). Both PPVT and NDVResid predicted children's accuracy on 3SG –*s*; the model with PPVT and NDVResid was more accurate than the model with only NDV as a predictor for lexicon size ($\chi 2 = 4.1447, df = 1, p < .05$). Recall that significant correlations also emerged between MOE and PPVT and between MOE and NDV, but the correlation between MOE and NDVResid was not significant ($r(67) = 0.07, p = .57$); hence, effects of MOE and NDVResid exist independent of each other. In order to verify if PPVT explained variation in the data when teased apart from MOE, we entered a decorrelated predictor PPVTResid in the model. The correlation between PPVT and PPVTResid was significant ($r(67) = 0.84, p < .0001$). Rerunning the model with PPVTResid revealed a separate effect for receptive vocabulary. We did not find any significant main or interaction effects for the expressive vocabulary measure *D* or for word and lemma frequency in the COCA corpus.

The final Model 1 (based on 1,125 observations, 196 verbs, and 15 children) is summarized in Table 4. The model performs well: $C = 0.93$. The effects of word frequency, NDV, PPVT, MOE as well as the interactions between lemma frequency and L1 and between Allomorph and L1 are plotted in Figures 2a–f. The plots show partial effects, that is, the effect of a factor while all other factors are calibrated for the reference level and continuous factors for the median. For the decorrelated word frequency predictor in the Edmonton ELL corpus, the median is 0.15, for lemma frequency in the Edmonton ELL corpus 5.48, for NDVResid -1.28, for PPVTResid 5.7, and for MOE 25. The reference level of Allomorph is /ɪz/ and the reference level of L1 is isolating.

In Figures 2a–f, the log odds ratios are back-transformed in proportions, but because the proportions correct represent partial effects they do not provide information on mean proportions correct in the whole sample, in contrast to Figures 1a–c. The plots in Figures 2a–f are important for information about the slope of the curves (Figures 2a–e) and the estimated difference between levels of a predictor (Figure 2f), and they are particularly useful for interpreting

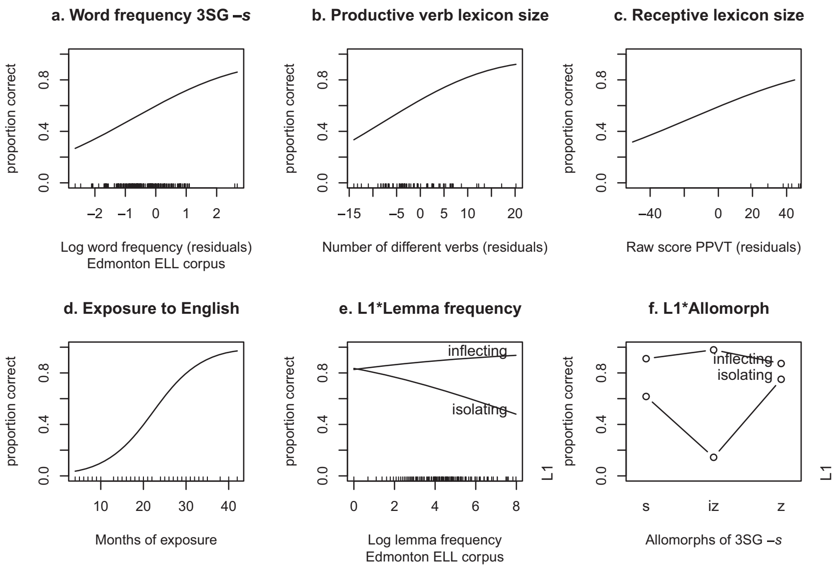
Table 4 Model estimates, standard error (*SE*), *Z*-values, and *p*-values of Model 1

	Coefficient	*SE*	*Z*	*p*
(Intercept)	−6.32	1.04	−6.09	$p < .0001$
Word frequency 3SG –*s* Edmonton ELL corpus-residuals	0.53	0.16	3.49	$p = .0005$
Lemma frequency Edmonton ELL corpus	−0.21	0.08	−2.61	$p = .0091$
Allomorph (/s/)	2.26	0.63	3.60	$p = .0003$
Allomorph (/z/)	2.88	0.62	4.63	$p < .0001$
NDV-residuals	0.09	0.02	4.43	$p < .0001$
PPVT-residuals	0.02	0.01	2.16	$p = .0312$
L1 (inflecting)	3.73	1.62	2.31	$p = .021$
MOE	0.14	0.02	5.80	$p < .0001$
L1 (inflecting)*Lemma Frequency Edmonton ELL corpus	0.36	0.11	3.30	$p = .001$
L1 (inflecting)*Allomorph (/s/)	−3.80	1.28	−2.96	$p = .003$
L1 (inflecting)*Allomorph (/z/)	−4.81	1.28	−3.77	$p = .0002$

Note. NDV = Number of different verbs; PPVT = Peabody Picture Vocabulary Test; MOE = Exposure to English in months. *Interaction effect.

interaction effects (Figures 2e–f). The coefficients are all simultaneously present in the model. Consequently, if for a certain predictor both significant interaction and main effects emerge, the coefficient for the main effect does not represent the simple main effect familiar from sequential analysis of variance tables. In such cases, we focus on the interaction effects.

The ascending curves in Figures 2a, 2b, 2c, and 2d illustrate that the children performed more accurately with verbs they heard more often with 3SG –*s* in the input ($p < .001$), that children with a larger productive verb lexicon ($p < .0001$) and children with larger receptive lexicons were more accurate with 3SG –*s* ($p < .05$), and that the children improved their accuracy with 3SG –*s* after more months of exposure to English ($p < .0001$). Children with an isolating L1 performed worse on 3SG –*s* with verbs that have higher lemma frequency in the Edmonton ELL corpus ($p < .001$). Finally, children with isolating L1s performed better with both /s/ and /z/ than with /ɪz/ ($p < .001$; $p < .001$). After altering the reference level of Allomorph to /s/, we found that children with an isolating L1 were more accurate with the



Figures 2 a–f Main and interaction effects in model 1

allomorph /z/ than with /s/ ($p < .05$). Children with inflecting L1s did not show this difference between the allomorphs.

In order to take a closer look at the observed effects of input distributions, three follow-up analyses were performed. First of all, we tested if word frequency and lemma frequency could be reduced to one effect: relative word frequency. Second, it was examined if a specific component of lemma frequency caused the lower performance in the isolating L1 group. Third, we investigated frequency distributions of the allomorphs with the purpose to determine if input frequencies could have caused the allomorph effect in isolating L1 children. We present each set of results in turn.

Relative Word Frequency

It is conceivable that word frequency and lemma frequency (which comprises bare verb forms, verbs ending on 3SG –*s*, past tense forms and participles) could be reducible to one effect. This relative word frequency would mean that the children perform accurately with those verbs that appear often in 3SG form and infrequently appear as bare, past or participle forms, and they may perform worse on verbs with a reverse distributional pattern. In the latter case, the frequency of forms other than 3SG –*s* for certain verbs may inhibit the lexical retrieval of [verb+3SG –*s*] in obligatory contexts, as explained in our review

Table 5 Model estimates, standard error (*SE*), *Z*-values, and *p*-values of Model 3

	Coefficient	*SE*	*Z*	*p*
(Intercept)	−6.65	1.20	−5.56	$p < .0001$
Word frequency 3SG –*s* Edmonton ELL corpus	0.52	0.18	2.87	$p = .0041$
Word frequency bare verb Edmonton ELL corpus	−0.56	0.14	−3.94	$p < .0001$
Allomorph (/s/)	2.27	0.67	3.37	$p = .0008$
Allomorph (/z/)	2.89	0.67	4.32	$p < .0001$
NDV-residuals	0.09	0.03	3.12	$p = .0019$
PPVT-residuals	0.03	0.01	2.35	$p = .0189$
MOE	0.13	0.03	4.25	$p < .0001$

Note. NDV = Number of different verbs; PPVT = Peabody Picture Vocabulary Test; MOE = Exposure to English in months.

of the literature. In order to test relative frequency, we ran an alternative model (Model 2, not shown in this article) that contained a factor which estimated for each verb the proportion of 3SG –*s* forms instead of two separate factors, as in Model 1. A comparison of Models 1 and 2 indicated that the more complex Model 1 in which word frequency and lemma frequency were separated was still preferred, albeit hardly ($\chi 2 = 3.8288$, $df = 1$, $p = .05$).

Lemma Frequency Decomposed

It was found that a higher lemma frequency for a verb was associated with more omissions of 3SG –*s* with that verb in children with an isolating L1. Lemma frequency is comprised of five word frequencies: frequency of 3SG –*s*, bare verbs, past tense verbs, past participles, and progressive participles. Exploring the hypothesis that particular forms are responsible for the effect of lemma frequency, we broke lemma frequency down into separate frequencies of bare forms, past tense forms, past participles, and progressive participles and then regressed isolating L1 children's performance on 3SG –*s,* entering these separate predictors instead of the combined predictor lemma frequency.

It turned out that the more frequently a verb appeared in bare form, the more likely it was that a child omitted 3SG –*s* ($p < .0001$). The frequency of past tense forms indicated a trend but was not significant at the alpha level of .05, and the model without this predictor was still preferred ($\chi 2 = 3.4287$, $df = 1$, $p = .06$). The outcomes of this model, Model 3, are shown in Table 5.

Table 6 Distributions of 3SG –*s* allomorphs /ɪz/, /s/, and /z/

	Corpus	*k* overall[a]	*k* verb types[b]	Word frequency		
				Min-max	*Mean*	*SD*
/ɪz/	Edmonton ELL	71(2%)	23(10%)	0–15	4	5
	COCA	32, 010(6%)	738(16%)	0–1,713	517	543
/s/	Edmonton ELL	735(25%)	83(35%)	0–126	46	43
	COCA	192, 061(36%)	1, 753(38%)	0–18,322	7,376	6,853
/z/	Edmonton ELL	2, 156(73%)	129(55%)	0–552	54	82
	COCA	309, 432(58%)	2, 121(46%)	0–21,073	8,553	8,188

[a]For each allomorph, raw total token count (and percentage of all 3SG verb tokens in corpus).
[b]For each allomorph, raw total type count (and percentage of all 3SG verb types in corpus).

It is based on 621 observations, 145 verbs, and 9 children. The model fit is good ($C = 0.94$).

Allomorph Frequencies in the Input

Model 1 indicated that children with isolating L1s used /z/ correctly most often and that particularly /ɪz/ was prone to errors. Because developmental variation between allomorphs may be related to frequency distributions in the input, we compared distributional properties of the three allomorphs; the relevant information is listed in Table 6. The table gives overall frequency information, that is, how often an allomorph is used, broken down into detailed information about the number of different verbs with which the allomorph was used (i.e., type frequencies), and how frequent individual verbs with a certain allomorph were (i.e., word frequencies).

In both corpora, /ɪz/ had the lowest frequency and appeared with the lowest number of verb types, whereas /z/ was most frequent and was used with most verb types; /s/ was in between.[1] Also, there were no verbs with high word frequencies in this dataset that take /ɪz/, and there were more verbs with high word frequencies that take /z/ than /s/. As shown in Table 6, the frequency data are consistent across the two corpora.

Discussion

The results of our study can be summarized in the following four main findings: (i) English L2 children were more accurate with verbs with a high word frequency of 3SG –*s*; (ii) children with isolating L1s were less accurate than children with an inflecting L1, in particular with verbs that had a high lemma frequency in the input, and verbs that take the allomorph /ɪz/, and to a lesser extent the allomorph /s/; (iii) a larger lexicon was associated with better performance on 3SG –*s*, and this effect was found for both productive verb lexicon and for receptive lexicon; and (iv) the children were more accurate after more months of exposure to English. Follow-up analyses additionally revealed that (v) the negative effect of lemma frequency was caused by the word frequency of bare forms and (vi) the allomorph effect lined up with frequency distributions in the input.

This study was designed to explore factors that impact on the development of 3SG –*s* in child L2 English. Our specific goal was to answer four research questions about the role of input frequencies, allomorphs, lexicon size, and L1 on L2 children's development of 3SG –*s*. Below we summarize the contribution of these factors in our study, and discuss our findings in the light of results of other studies and theoretical approaches to the learning of inflection, in particular Bybee's (1995, 2001, 2008, 2010) usage-based NM.

Frequency Effects

The effect of word frequency confirms predictions of the NM, and is consistent with previous experimental findings from English L1 children which showed a relationship between the form in which children had learned novel verbs in 3SG contexts and how they produced them (Finneran & Leonard, 2010; Theakston et al., 2003). Song et al. (2009) analyzed naturalistic input data, but did not find any correlations between input distributions and English L1 children's production of 3SG –*s*. It could be that frequency data in the CHILDES database consulted are not representative of the input of the six individual children studied by Song et al. (2009). Another possibility is that phonotactic probability of consonant clusters has greater predictive value in L1 acquisition than word frequency. In the latter case, the observed discrepancy of findings across studies could mean that whereas phonotactic probabilities in the input impact the omissions of 3SG –*s* in younger (L1) children, distributional properties with higher granularity, such as word frequency, influence omissions in older (L2) children.

We found that children with an isolating L1 dropped 3SG –*s* more often with verbs that appear frequently in bare form in the input. The impact of other verb forms on a particular verb form is compatible with the NM, because, according to this theory, each word is part of a network of associated words based on semantic and phonological similarities between these words. If for a given verb the bare form has a stronger lexical representation than the inflected form, a child might access the bare form more easily than the inflected form, resulting in errors in 3SG contexts. The impact of the frequency of other paradigmatic forms on lexical access of a specific form is consistent with the outcomes of lexical decision experiments conducted with adults showing similar effects (Milin, Kuperman, Kostić, & Baayen, 2009; Moscoso del Prado Martín, Kostić, & Baayen, 2004).

In previous studies on the learning of inflection, effects of word and lemma frequency have been interpreted in the light of dual versus single mechanism views (Pinker & Ullman, 2002). The NM, a single mechanism view, predicts that word frequency impacts on children's accuracy with regularly inflected verbs, such as 3SG –*s*, because these verbs are stored as whole forms. Each encounter with this inflected form will make a child's lexical representation of this form stronger and less prone to errors in production. According to a dual mechanism view, verb stem and inflectional affix are stored separately in the lexicon, and inflected words are computed online using categorical grammatical rules, which are not part of the lexicon (Marcus et al., 1992). This view may predict effects of lemma frequency, because lemma frequency affects retrieval of the stem, but no effects of word frequency are expected (Baayen, Schreuder, De Jong, & Krott, 2002), or only to a very limited extent (Prasada & Pinker, 1993). The effect of word frequency in our study is therefore difficult to reconcile with a dual mechanism view of inflectional learning.

In order to estimate effects of word frequency and lemma frequency, two corpora were consulted. The first corpus was a large spoken language corpus of North-American English, mainly based on the language used in television shows (COCA). The second corpus was a small corpus of Western-Canadian English with language used by adults and English L2 children during a play session (the Edmonton ELL corpus). The effects of word frequency and lemma frequency emerged only in the latter corpus, notwithstanding the small corpus size. The different outcomes for the two corpora illustrate that not all spoken language corpora are representative of the input that a certain learner group is exposed to, and that this effect cannot be overcome by corpus size.

Variation Between Allomorphs

Isolating L1 children's tendency to drop /ɪz/ more often than the other allomorphs of 3SG –*s* parallels observations for plural /ɪz/ and past tense /ɪd/ in English L1 children (Berko, 1958; Köpcke, 1998; Bybee, 2007). Vulnerability of 3SG /ɪz/ may have several, partially related, reasons. First, L2 children may have dropped /ɪz/ because they rely on a schema that states that verbs ending in /s/ or /z/ express 3SG. This explanation has also been offered in the aforementioned studies on L1 English. Second, /ɪz/ may have a weak schema due to low type frequencies of this allomorph. According to Bybee (2001), a low type frequency will make it more difficult to identify (allomorphs of) suffixes, and schemas for such endings will be weak. Third, given that /ɪz/ will not be productive due to a weak schema, its emerging correct use will be dependent on word frequency. In the corpora we consulted, word frequencies of most verbs on /ɪz/ were considerably lower than word frequencies of most verbs on /s/ or /z/, and high word frequencies were only found for /s/ and /z/, and not for /ɪz/.

The difference between 3SG /s/ and /z/ in our study resembled Köpcke's (1998) observation that in child L1 English performance with plural /s/ is less accurate than with /z/. The difference between /s/ and /z/ in our study lines up with input distributions of these allomorphs, as the type frequency of /z/ exceeded the type frequency of /s/ and word frequencies for /z/ were higher than for /s/. Given these frequency distributions, the NM would predict that the /z/ schema is more productive and stronger than the /s/ schema, and therefore is more likely to be successfully generalized by children to less frequent verbs that would fit this schema than the /s/ schema. Also, because word frequencies for 3SG verbs ending in /z/ are generally higher than for verbs ending in /s/ the overall greater accuracy with /z/ could be higher due to the lexical strength of individual 3SG verbs.[2]

Lexicon and Grammar

The effect of (productive verb and receptive) lexicon size on L2 children's accuracy with 3SG –*s* ties in with findings about monolingual children (Bates et al., 1994; Caselli et al., 1999; Fenson et al., 1994), bilingual toddlers (Conboy & Thal, 2006; Marchman et al., 2004), and preschool and school-aged L2 children (Cobo-Lewis et al., 2002; Simon-Cereijido & Gutiérrez-Clellen, 2009), all showing that lexical and grammatical development are related. The outcomes of Model 1 indicated that the specific measure of verb lexicon size was a more important predictor than the general measure of receptive vocabulary PPVT. This is consistent with other studies that report correlations between closely related lexicon and grammar measures (Marchman & Bates, 1994; Simon-Cereijido & Gutiérrez-Clellen, 2009). Using mixed regression models, we ensured that the correlation between lexicon size and accuracy of 3SG –*s* was assessed within individual children. This method of correlating related lexical and grammar measures within individual children is desirable for addressing questions about their relationship, yet it is rarely applied in published studies (Rispoli, Hadley, & Holt, 2009).

L1 Transfer

Children in the two L1 groups had comparable levels of English proficiency, as measured by receptive vocabulary, expressive vocabulary, and grammar. The lower performance of the isolating L1 children on 3SG –*s* is, therefore, unlikely to be part of an overall delay in L2 English, but indicates effects of L1 transfer in the domain of inflection. This effect of transfer supports other studies that report L1 effects in English L2 children's acquisition of verb inflection (Dulay & Burt, 1974; McDonald, 2000; Paradis, 2011). The children in our study also participated in other studies where they showed effects of transfer in learning the English article (Zdorenko & Paradis, 2008, 2012) and auxiliary (Zdorenko, 2010) systems; they thus consistently transfer L1 properties across different grammatical morphemes. Paradis (2005) and Paradis et al. (2008) did not find differences in accuracy on verb inflection between children whose L1 was isolating and children whose L1 was richly inflecting. Although the children in these studies are from the same sample as the children in our study, they were studied at round 1 only, and floor effects in children's accuracy might have been the cause for the absence of between-group differences. For example, Paradis (2005, p. 180) found that children were 18% accurate with 3SG –*s* in spontaneous language samples, and 16% accurate with 3SG –*s* on an elicitation probe.

In our review of the literature, we distinguished between transfer at the level of representations, that is, L1 schemas may feed into L2 schemas based on overlap (Gathercole, 2007; MacWhinney, 2008a), and transfer at the level of perceptual focus of attention, which may be shaped by the L1 (Ellis, 2006, 2008). In both cases inflecting L1 children will develop their knowledge of 3SG –*s* faster than isolating L1 children. Isolating L1 children showed a tendency to drop 3SG –*s* with verbs that often appear in bare form in the input. The absence of such an effect for inflecting L1 children suggests that for them retrieval of a verb form is less dependent on frequency information on the word level and, at the same time, more dependent on contextual cues prompting retrieval of the correct inflectional form given a 3SG subject. Thus, the same input cues (e.g., subject, word frequencies) affect English L2 children's use of 3SG –*s* differently; which cue wins seems dependent on whether or not in the L1 subject and verb are associated. The competition model put forward by MacWhinney (2008a) appears to be consistent with this interpretation. The allomorph effects may indicate that isolating L1 children need more time to generalize over allomorphs that express the same semantic categories, namely 3SG, and are more sensitive to phonological detail in the input than inflecting L1 children. This would be expected given that isolating L1 children are fully dependent on input properties, similar to English L1 children who also show allomorph effects (Berko, 1958; Köpcke, 1998; see also Bybee, 2007). Inflecting L1 children, in contrast, can make use of prior knowledge that provides them with the ability to quickly expand schemas for 3SG –*s* and generalize over the allomorphs of this suffix.

Conclusion

The development of 3SG –*s* in the English L2 children who participated in our study was related to word frequency, lexicon size, and months of exposure to English. Accuracy with 3SG –*s* in children with an isolating L1 varied across allomorphs and was determined by lemma frequency in the input, in particular by the frequency of bare verbs, a component of lemma frequency. According to the NM (Bybee 1995, 2001, 2008, 2010), word frequency and amount of exposure will be important because this model predicts that frequency of use determines language development; lexicon size may predict L2 children's accuracy with 3SG –*s* because children with larger vocabularies will have stronger schemas for 3SG –*s*. Isolating L1 children lack prior knowledge that facilitates generalizations over 3SG –*s*. Consequently, they distinguish between allomorphs, like English L1 children. If a verb is often bare, retrieval of 3SG –*s* for this verb may be inhibited. This effect emerged in isolating L1 children only because these children may attend less to contextual cues that are relevant for selecting 3SG –*s* than inflecting L1 children. We conclude that the results of this study emphasize that a range of factors simultaneously impact language development, and indicate that a usage-based approach, such as the NM, can successfully account for the factors that determine L2 children's accuracy with using verb inflection.

Revised version accepted 11 April 2011

Notes

1 As pointed out by one reviewer, Song et al.'s (2009) observation that English L1 children drop 3SG –*s* less often with singleton clusters could create a confound between allomorph and phonological complexity because verbs that end in a vowel all take the allomorph /z/. However, coda complexity did not emerge as a

significant predictor in the preliminary analyses for the present study, unlike differences between allomorphs.

2 Köpcke (1998) argues that because more singular nouns end with /s/ than /z/ the distance between /s/ and the canonical plural marker will be greater than for /z/ and the canonical plural marker. In addition to proximity, Köpcke mentions (perceptual) salience and iconicity as relevant factors. These factors may also have contributed to allomorph effects in our study.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Schreuder, R., De Jong, N. H., & Krott, A. (2002). Dutch inflection: The rules that prove the exception. In S. Nooteboom, F. Weerman, & F. Wijnen (Eds.), *Storage and computation in the language faculty* (pp. 61–92). Dordrecht, Netherlands: Kluwer.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P. S., & Reznick, J. S. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, *21*, 85–123.

Berko, J. (1958). The child's learning of English morphology, *Word*, *14*, 150–177.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Butt, J., & Benjamin, D. (2004). *A new reference grammar of modern Spanish* (4th ed.). London: Hodder Education.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455.

Bybee, J. (2001). *Phonology and language use*. Cambridge, UK: Cambridge University Press.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford, UK: Oxford University Press.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). New York: Routledge.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge, UK: Cambridge University Press.

Caselli, M. C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*, *26*, 69–111.

Chatterjee, C., & Hadi, A. S. (2006). *Regression analysis by example* (4th ed.). New York: Wiley.

Cobo-Lewis, A. B., Eilers, R. E., Pearson, B. Z., & Umbel, V. C. (2002). Interdependence of Spanish and English knowledge in language and literacy among bilingual children. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 118–132). Clevedon, UK: Multilingual Matters.

Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development*, *77*, 712–735.

Dulay, H., & Burt, M. (1973). Should we teach children syntax? *Language Learning*, *23*, 95–123.

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

Dunn, L. Dunn, L. (1997). *The Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking and perceptual learning. *Applied Linguistics*, *27*, 164–194.

Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associate learning of constructions, learned attention, and limited L2 endstate. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372–405). New York: Routledge.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), 1–185 [Serial No. 242].

Finneran, D. A., & Leonard, L. B. (2010). Role of linguistic input in third person singular *–s* use in the speech of young children. *Journal of Speech, Language and Hearing Research*, *53*, 1065–1074.

Gathercole, V. C. M. (2007). Miami and North Wales, so far and yet so near: Constructivist account of morpho-syntactic development in bilingual children. *International Journal of Bilingual Education and Bilingualism*, *10*, 224–247.

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority L1 children learning English as L2. *Applied Psycholinguistics*, *29*, 1–25.

Gönczöl-Davies, R. (2007). *Romanian: An essential grammar*. New York: Routledge.

Haznedar, B. (2001). The acquisition of the IP system in child L2 English. *Studies in Second Language Acquisition*, *23*, 1–39.

Ionin, T., & Wexler, K. (2002). Why is "is" easier than "s"? Acquisition of tense/agreement morphology by child L2 English learners. *Second Language Research*, *18*, 95–136.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59, 434–446.

Kohnert, K., Pui Fong, K., & Conboy, B. (2010). Lexical and grammatical associations in secquential bilingual preschoolers. *Journal of Speech, Language and Hearing Research*, *53*, 684–698.

Köpcke, K-M. (1998). The acquisition of plural marking in English and German revisited: Schemata vs. rules. *Journal of Child Language*, *25*, 293–319.
Lin, H. (2001). *A grammar of Mandarin Chinese*. Munich, Germany: Lincom Europa.
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum.
MacWhinney, B. (2008a). Usage-based and form-focused language acquisition: The associate learning of constructions, learned attention, and limited L2 endstate. In P. Robinson & N. Ellis, (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 341–371). New York: Routledge.
MacWhinney, B. (2008b). Enriching CHILDES for morphosyntactic analysis. In H. Behrens (Ed.), *Trends in corpus research: Finding structure in data* (pp. 165–197). Amsterdam: Benjamins.
Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*, 85–104.
Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, *12*, 339–366.
Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science*, *7*, 212–224.
Marcus, G., Pinker, S., Ullman, M., Hollander, J., Rosen, T., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, 57(4), 1–165 [Serial No. 228].
Matthews, S., & Yip, V. (1991). *Cantonese: A comprehensive grammar*. London: Routledge.
McDonald, J. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, *21*, 395–423.
Milin, P., Kuperman, V., Kostić, A., & Baayen, R. H. (2009). Paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 214–252). Oxford, UK: Oxford University Press.
Moscoso del Prado Martín, F., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information-theoretical perspective on morphological processing. *Cognition*, *94*, 1–18.
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 the Schools*, *36*, 172–187.
Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*, 213–237.

Paradis, J., Rice, M., Crago, M., & Marquis, J. (2008). The acquisition of tense in English: Distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics*, *29*, 689–722.
Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Science*, *6*, 456–463.
Prasada, S., & Pinker, S. (1993). Generalizations of regular and irregular morphology. *Language and Cognitive Processes*, *8*, 1–56.
Rispoli, M., Hadley, P., & Holt, J. (2009). The growth of tense productivity. *Journal of Speech, Language and Hearing Research*, *52*, 930–944.
Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, *30*, 315–337.
Song, J. Y., Sundara, M., & Demuth, K. (2009). Phonological constraints on children's production of English third person singular *–s*. *Journal of Speech, Language and Hearing Research*, *52*, 623–642.
Theakston, A. L., Lieven, E. V., & Tomasello, M. (2003). The role of input in the acquisition of third-person singular verbs in English. *Journal of Speech, Language and Hearing Research*, *46*, 863–877.
Tyler, A. (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics*, *30*, 270–291.
Zdorenko, T. (2010). *Constructions in child second language acquisition: Exploring the role of first language and usage*. Unpublished doctoral dissertation, University of Alberta.
Zdorenko, T., & Paradis, J. (2008). The acquisition of articles in child second language English: Fluctuation, transfer or both? *Second Language Research*, *24*, 227–250.
Zdorenko, T., & Paradis, J. (2012). Articles in child L2 English: When L1 and L2 acquisition meet at the interface. *First Language*, *32*, 38–62.