The effect of frequency and phonological neighbourhood density on the acquisition of past tense verbs by Finnish children*

MINNA KIRJAVAINEN, ALEXANDRE NIKOLAEV and EVAN KIDD

Abstract

The acquisition of the past tense has received substantial attention in the psycholinguistics literature, yet most studies report data from English or closely related Indo-European languages. We report on a past tense elicitation study on 136 4–6-year-old children that were acquiring a highly inflected Finno-Ugric (Uralic) language—Finnish. The children were tested on real and novel verbs (N = 120) exhibiting (1) productive, (2) semi-productive, or (3) non-productive inflectional processes manipulated for frequency and phonological neighbourhood density (PND). We found that Finnish children are sensitive to lemma/base frequency and PND when processing inflected words, suggesting that even though children were using suffixation processes, they were also paying attention to the item level properties of the past tense verbs. This paper contributes to the growing body of research suggesting a single analogical/associative mechanism is sufficient in processing both productive (i.e., regular-like) and non-productive (i.e., irregular-like) words. We argue that seemingly rule-like elements in inflectional morphology are an emergent property of the lexicon.

Keywords: *Language acquisition; Finnish; Inflectional morphology; Past tense.*

* We would like to thank the children and nurseries that took part in our study, and the three anonymous reviewers for their helpful comments. This research was supported by an Economic and Social Research Council grant to the third author (RES–000–22–2640). Address for correspondence: Minna Kirjavainen, School of Psychological Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK. Email: ⟨minna.kirjavainen@manchester.ac.uk⟩.

Cognitive Linguistics 23–2 (2012), 273–315
DOI 10.1515/cog-2012-0009

0936–5907/12/0023–0273

1. Introduction

The acquisition and processing of morphologically complex words (e.g., *jump+ed* [1]) has been the basis of a long and enduring debate in cognitive science (e.g., Albright and Hayes 2003; Berko 1958; Bybee 1988, 1995; Bybee and Slobin 1982; Marcus et al. 1995; Pinker 1991; Pinker and Prince 1988; Pinker and Ullman 2002). The disagreement centres on the question as to whether or not (1) speakers use different processes to form different classes of verbs, ostensibly 'regular' (e.g., *jump → jumped*) and 'irregular' (e.g., *say → said*) forms, or whether (2) all forms are processed by the same mechanism. The disagreement derives largely from different theoretical assumptions about language/language acquisition. The nativist stance posits innate grammatical categories and rule application, combined with memorised exceptions to the rules. In contrast, the usage-based approach assumes that children memorise and then schematise over stored items, a process that does not make a sharp distinction between regular and irregular forms. In this paper we will investigate whether data from a highly inflected language, Finnish, supports the former or latter stance.

Many languages have at least one 'regular' (also referred to as 'productive' or 'default') way to create a morphologically complex word (e.g., add an *-ed* suffix for past tense in English). In addition to these highly productive inflectional processes, languages can also have exceptions to these apparent rule-like operations (e.g., *sing → sang*, *think → thought*), also referred to as 'non-productive', 'irregular' and 'non-default' inflections. These seemingly distinct categories of lexical items, one of which exhibits relatively clear rule-like behaviour and the other one less so, has resulted in a processing model whereby the two groups are also assumed to be processed differently (e.g., Marcus et al. 1995; Pinker 1991; Pinker and Ullman 2002; Ullman 1999, 2001). The key assumption of the *dual route* approach is that regular and irregular morphological processes are qualitatively different: irregular morphologically complex items are processed by associative memory, whereas regular items are processed by a separate computational mechanism in which symbolic rules are applied to members of abstract categories (e.g., Verb, Noun). However, if putatively regular forms are high in frequency or are phonologically similar to irregular items, they can also be stored in associative memory (e.g., Hartshorne and Ullman 2006: 25; Ullman 2001: 53).

The dual route mechanism has been argued to operate in the following way. When producing inflected words, both the computational and associative memory routes are activated, but associative memory is searched for a whole

1. In this article, within examples a hyphen is used to highlight the syllable boundary and + a morpheme boundary.

word representation first (e.g., Pinker 1991; Ullman 2001). If a whole word representation is found (e.g., *went*), the computational route (e.g., **goed*) will be blocked and the stored item produced. If the associative memory does not contain a whole word representation (and hence block the use of a computational mechanism), a 'default' (i.e., regular) affix (e.g., *-ed*) is added to the base form of the word (e.g., *jump → jump+ed*) as a 'last resort' (Marcus et al. 1995: 238). Overgeneralization errors (i.e., an irregular item is inflected as if it was a regular item e.g., **goed*) are explained by speakers failing to find the irregular form in the associative memory. On the other hand, irregularization errors, in which a regular word is inflected as if it were an irregular item (e.g., *beep → *bept*), are thought to be rare and will only occur if a regular item rhymes with an irregular one, allowing irregularization via analogy in associative memory (Pinker 1991; Xu and Pinker 1995).

The application of the default affix is argued to be indifferent to factors such as phonological form of the word, type frequency, and, excluding highly frequent words, token frequency (e.g., Marcus et al. 1995; Pinker 1991; Ullman 2001). The default rule can be applied to any word that does not have the morphologically complex representation in associative memory. Hence, all regular words should be equally easy to produce provided that their base/lemma frequencies are similar. However, the frequency and phonological form of words are assumed to have an effect on the processing of irregular (i.e., non-default) words because they are stored in associative memory, and so are subject to the same constraints on lexical access as are morphologically simple words. The dual route model predicts that novel items are treated as regular items, unless they resemble an irregular word, in which case the past form will be predicted by analogical processes.

The single route approaches such as the schema (Bybee 1988, 1995; Bybee and Slobin 1982; Langacker 1987, 2000) and connectionist (e.g., Eddington 2009; Plunkett and Marchman 1993; Plunkett and Nakisa 1997; Rumelhart and McClelland 1986) accounts fundamentally differ from the dual route approach. Most importantly, these accounts assume no qualitative distinction between regular and irregular items. Instead, all morphologically complex words are thought to be processed in the associative memory by a single mechanism. According to the schema model, once a language learner has reached a certain threshold or a 'critical mass' (Marchman and Bates 1994) of whole (i.e., inflected) words stored in associative memory, schema-based abstractions will begin to be made on the basis of phonetic and semantic similarities between different stored items. When similarities are identified, related words are linked. These connections do not need to occur only at the word level (e.g., *ring → rang, sing → sang*), but can also occur at the bound morphemic (e.g., *-ed*) and phonemic level (e.g., /k/ in ***c**at*, ***c**aps* and ***k**itten*). The detection of these similarities result in the construction of schemas of varying abstraction

that simulate rule-like behaviour (e.g., CV*st*: *kissed, missed, hissed, guessed, fussed*, Dąbrowska 2008a). The single route approach therefore identifies a prominent role for item-level features in the construction of schema-based knowledge (Marchman 1997).

Within the single route account, the phonological form of a word in relation to other words the child knows is assumed to affect the processing of that word (Bybee 1995; Marchman 1997). If a word's base and inflected forms rhyme (e.g., *sing → sang, ring → rang*) with many other words (i.e., it has many phonological 'friends'), and if it has few/no 'enemies' (i.e., words whose base form rhymes with it but whose inflected form does not, e.g., *brake → braked*, *take → took*), the word should be easier to acquire/process than a word that has few friends and/or many enemies. There is some support for this assertion from children acquiring English, where children have been shown to acquire morphologically complex/simplex words that have many phonological friends earlier than words that have no/few friends (Marchman 1997; Storkel 2004). Similarly, overgeneralization errors are commonly caused by phonological similarity between the overgeneralized/irregularized verb and a group of verbs rhyming with it (Kidd and Lum 2008; Marchman 1997; Ragnarsdóttir et al. 1999).

Type and token frequencies are also assumed to affect the processing of both regular and irregular words (e.g., Bybee 1995; Marchman 1997). A high number of verb types contributing to a particular schema (e.g., the number of verbs taking the *-ed* past tense suffix in English in comparison to the type frequency of any particular sub-type of irregular inflections) strengthens that schema and results in greater ease of generalization (Bybee and Newman 1995; Dąbrowska 2004). This explains why overgeneralizations with the past tense *-ed* in English are commonplace while irregularizations are not. In addition, type frequency has been shown to have a facilitating effect on children's correct production of regular and irregular inflected items (Dąbrowska 2004; Ragnarsdóttir et al. 1999).

Inflected words with high token frequency are predicted to be acquired earlier; since these forms are encountered frequently they are likely to be stored and incorporated into the existing representations. On the other hand, words with low token frequency are assumed to exhibit higher error rates. Studies on children from a variety of language backgrounds have shown that children are better at inflecting high than low token frequency regular and irregular words (e.g., Dąbrowska 2001, 2004; Marchman 1997; Ragnarsdóttir et al. 1999).

Although child language data provides some evidence for frequency and PND effects in the production of regular/irregular words, such effects have yet to be established in a language such as Finnish, where a rich system of verbal morphology means that any one verb has potentially thousands of inflected forms (Karlsson and Koskenniemi 1985). In the following section we describe Finnish past tense morphology.

1.1. *Finnish*

Finnish verbs (N = 9265 excluding compound verbs [Suomen kielen perussanakirja 1994]) can be categorised into 25 verb classes (ibid.) based on their (morpho)phonology and the way they inflect. Verbs are inflected for four forms of infinitive, two forms of participle, four forms of mood, present and past tense, active and passive, and six forms of person (1st, 2nd, 3rd person singular and plural). In addition, verbs can take clitic-like suffixes (e.g., kO, hAn, pA[2]). Affixes denoting these inflections can be, and commonly are, added to the stem of the verb. See example (1).

(1) *pela* *+si* *+tte* *+ko*
play +PAST +1pl +interrogative clitic
"Did you play?"

The high number of verbal inflections, clitics, and the fact that one or more morphemes can be added to the stem results in any given Finnish verb being able to occur in over 10 000 different forms (Karlsson and Koskenniemi 1985). However, if extra-paradigmatic clitic-like suffixes are excluded, the number of forms in which a given verb is likely to appear is somewhere around 260 (Hakulinen et al. 2005: 85), which nevertheless presents a challenging task for a language learner (cf. English, which has four or five different forms for each verb).

Finnish has two past tense suffixes: *-si* and *-i* (Hakulinen et al. 2005: 91, 140; Karlsson 1977; Suomen kielen perussanakirja 1994; see also Niemi 2006; Niemi and Niemi 1987; Räisänen 1975).[3] In colloquial speech *-si* is often shortened to *-s* (e.g., *pela+s* instead of *pela+si* 'to play') (Hakulinen et al. 2005: 140), and *-i* may be omitted (e.g., *keinu+0* instead of *keinu+i* 'to swing'). Even though the English past tense allomorphs [d], [t] and [ɪd / əd] might tempt one to assume that the Finnish *-si* and *-i* are different phonetic realizations of the same suffix, this is not the case. The existence of two distinct past tense suffixes in Finnish is supported by the following points. First, a number of Finnish verbs take either *-si* or *-i* suffix in colloquial dialect varieties (e.g., *kaatua – kaatu+i* or *kaatu+si*[4] 'to fall over'), and in standard Finnish (e.g., *yltää – ylt+i – yl+si* 'reached', for more examples and a discussion see

2. Capital letters denote vowel harmony (in non-compound words) whereby the quality of a back or front vowel within the suffix can be affected by the quality of the vowel in the first syllable of the word. Only back and neutral vowels or front and neutral vowels can occur together in a given word.
3. Note that some other languages, for instance Hausa (e.g., Jaggar 2001), also have more than one past tense form.
4. 88% of the Kirjavainen-Max Planck Finnish corpus adults' 3sg past tense forms were *kaatusi,* the rest *kaatui.*

Hakulinen et al. 2005: 141–142). This illustrates that the two are not always in complementary distribution (unlike the different realizations of *-ed*). Second, the suffix plays a role in the selection of the stem form rather than the stem alone determining the suffix selection (see e.g., Cathey and Wheeler 1986; Hakulinen et al. 2005; Karlsson 1983). For instance, in words such as *ede+tä* ('to proceed') the past tense suffixation results in a different stem allomorph being selected *(eten+i)*. This type of a process contrasts with the English *-ed*, whose form, in most cases, is determined by the form of the stem. Finally, in spoken language the *-si* suffix is often realized as [s] (see the Discussion for more detail on the use of [s]), which makes the two suffixes in spoken language perceptually very different (*pela+s* vs. *nukku+i*|*nukku+0*), questioning whether young children, who are exposed to standard (written) language relatively infrequently, would ever treat the two as different realizations of a single suffix. To highlight the fact that *-si* and *-i* are not different realizations of the same suffix, in this paper we will refer to them as two past tense suffixes rather than two past tense allomorphs.

The *-si* suffix is used with so-called 'contracted verbs' that, when in the present indicative active forms have, for instance, a long single vowel sequence in the end of the word, as in (2), or a combination of vowels uA, eA, OA, iA, as in (3)–(6). On the other hand, *-i* is used, for instance, with verbs in which the labiality of the vowel in the first syllable of the stem results in the stem-final 'A' being deleted when preceding the past tense suffix, as in (7) and (8). A small number of verbs show zero past tense suffixation (9), and in four of the 25 verb categories the two suffixes can alternate relatively freely (10).

	A-INF	**3sg PRES**	**3sg PAST**	**GLOSS**
(2)	*pela+ta*	*pel**a+a***	*pela+si*	to play
(3)	*halu+ta*	*hal**u+a**a*	*halu+si*	to want
(4)	*nime+tä*	*nim**e**+**ä**ä*	*nime+si*	to name
(5)	*inho+ta*	*inh**o+a**a*	*inho+si*	to despise
(6)	*hävi+tä*	*häv**i**+**ä**ä*	*hävi+si*	to lose
(7)	*muista+a*	*muista+a*	*muist+i*	to remember
(8)	*näyttä+ä*	*näyttä+ä*	*näytt+i*	to show
(9)	*salli+a*	*salli+i*	*salli+0*	to allow
(10)	*ylt+ää*	*ylt+ää*	*yl(t)+si*|*i*	to reach
(11)	*pum**p**a+ta*	*pum**pp**a+a*	*pum**pp**a+si*	to pump
(12)	*pe**l**ä+tä*	*pe**lk**ä+ä*	*pe**lk**ä+si*	to be scared of

In many cases the two suffixes are attached without any changes to the stem. However, suffixation can and often does change the form of the stem. The *-i* suffixation can result in some changes to the end of the stem. These are nevertheless at least partly productive (we will return to this below). Also, some thirty percent of the verbs taking *-si* show consonant gradation[5], whereby a weak (short) consonant changes into a strong (geminate) consonant between the A-infinitive and past tense forms (Hakulinen et al. 2005) (e.g., *pum**p**ata* → *pum**pp**asi*). There are two types of consonant gradation changes: quantitative and qualitative. Quantitative changes, as in (11), result from a phonological process akin to the English past tense allomorph (t/d/ɪd/əd) process. Like the selection of the past tense allomorph in English, quantitative consonant gradation is a relatively productive process. On the other hand, in qualitative changes either a particular phoneme changes to another, or a phoneme is added to the stem (see example 12).[6] This process (like Finnish past tense inflections overall) is completely regular, but crucially it is not productive. For instance, novel words are unlikely to be inflected so that they show qualitative changes to the stem. Because of their very limited level of productivity, qualitative consonant gradation verbs can be seen as showing similarities to English irregular verbs. In this paper these will be referred to as *non-productive verbs*[7].

1.2. *Is there a default process in the Finnish past tense?*

Marcus et al. (1995) list 21 features that can be used to determine the default suffix for a particular inflection. However, this list was of little help to us in establishing a candidate default past tense suffix (*-si*, *-i*) in Finnish, just as Dąbrowska (2004) found in her work on Polish noun inflections (see also Ravid et al. 2008 in relation to Austrian German, Danish, Dutch and Hebrew noun plurals). For instance, many new loan words that enter Finnish take the *-si* suffix (Hakulinen et al. 2005; Martin 1989), suggesting that the *-si* is the default. However, problematically, *-i* can also be used for loan words (e.g.,

5. Consonant gradation = change in the length or quality of the consonant.
6. Note that it is not clear which stem form(s) children assume is the base form. For instance, Karlsson (1983) argues that the base is the 3sg present tense (indicative active) stem but our data suggests that the A-infinitive form (and possibly also the 3sg present tense form) affects the inflection of the word.
7. In this paper, different types of verbs in Finnish are not referred to as *regular*, *semi-regular* and *irregular* because all our verb types are 'regular', i.e., the formation of the inflected form is schema/rule-based as the phonological properties of the stem and the affix determine the inflection of the word. Instead, we will refer to the different verb types as *productive*, *semi-productive* and *non-productive* verbs, since they differ in their level of productivity/openness. Of course this does not mean that the verbs themselves are *productive*/*semi-productive*/*non-productive*, but rather that the verb class which they are an instance of is such.

allergistaa – allergisti 'to cause an allergy'). Both *-si* and *-i* can be used with denominalized verbs (e.g., *aita+si* 'fenced', *luonnostel+i* 'drafted'). This is relevant as words that have been derived from other categories must be headless and past tense information cannot be assigned to them in the lexicon (Marcus et al. 1995). That is, according to Marcus et al. (1995), these words must be processed by a default rule. Furthermore, both can be used with onomatopoeic words (e.g., *surra+si* 'buzzed', *liritt+i* 'liquid running'), which should only take the default suffix, and both are overgeneralized by children (e.g., Niemi 2006; Niemi and Niemi 1987; Räisänen 1975). Thus, based on Marcus et al.'s criteria, there is no one single default in Finnish.

In more general terms, Finnish past tense inflections (including both *-si* and *-i* suffixation) can be seen as highly predictable. The past tense inflection is determined by the combination of phonological properties of the stem and the suffix, and hence this process is totally 'regular' (e.g., Cathey and Wheeler 1986; Hakulinen et al. 2005). However, the *-si* and *-i* verb classes differ in their level of openness. *-si* suffix classes are more 'open' and consequently *-si* is used with most loan words. Because the *-si* suffix is the highly productive (i.e., is readily applied to unfamiliar words) in this paper we will refer to *-si* suffix verbs which do not show qualitative consonant gradation as *productive verbs*. As the *-i* suffix verb classes show some level of openness (as shown by the examples above), but not to the same degree as the *-si* suffix classes, we will refer to *-i* suffix verbs which do not show qualitative consonant gradation changes as *semi-productive verbs*. Finnish past tense inflection therefore does not consist of a clear-cut 'default' vs. 'non-default' process. Thus, we will not assign either suffix with a default status from the onset of the analysis. Instead, we will analyse the effects of frequency and PND on the production of *-si* and *-i* suffix verbs separately from each other and perform a detailed error analysis on both verb types to see whether the children's behaviour indicates default status for one or the other, or perhaps for both.

It might be worth noting that even though some proponents of the dual route model have acknowledged that it is possible to have more than one default (Pinker 1999; Pinker and Prince 1994), the default vs. non-default distinction has a central role in some dual route theories (e.g., Marcus et al. 1995; Ullman 2001, 2004). These latter scholars argue for a primary and fundamental role for rule-based processes in language. For instance, Ullman (2001: 43) makes the strong claim that only fully-productive morphological processes are rule-driven and hence become defaults. Hence, according to some dual route proponents, either *-si* or *-i* suffix verbs (or both) that show no qualitative consonant gradation changes (i.e., are 'regular-like' and at least semi-productive) should not show PND or token frequency effects, unless the target word rhymes with an irregular (non-productive verb) or has high token frequency—attributes that none of our test verbs had.

The predictions of the single and dual route approaches are as follows:

(1) The dual-route model
   (i) Either -*si* or -*i* (or both) should be the default affixation process and consequently should not show token frequency or PND effects.
   (ii) Non-productive (i.e., qualitative consonant gradation) verbs should show these effects.
   (iii) All novel verbs will be inflected using the default suffix (either -*si* or -*i,* or both if one assumes two default processes).
   (iv) Overgeneralization errors should be observed from the direction of the default suffixation process to the non-default, but not the other way around. However, if two default processes are assumed, overgeneralizations can be produced with either suffix. Note that if neither is the default overgeneralizations should be rare. This would indicate that the Finnish past tense inflectional process is totally associative memory based.

(2) The single-route model
   (i) All of our verb types should show frequency (when applicable) and PND effects.
   (ii) Overgeneralization errors can be produced with either past tense suffix if the child observes similarities between the target word and words that belong to a different suffix category.
   (iii) A higher type frequency of the -*i* suffix may result in overgeneralizations due to its higher generalisability.

1.3. *The current study*

The aim of our study was to test the dual and single route models' predictions about the processing of productive, semi-productive and non-productive words in Finnish. In particular, we investigated the effect of frequency and/or phonological neighbourhood density (PND) on the different verb types, and the error types the children produced.

2. Method

2.1. *Participants*

One-hundred-and-thirty-six (71 male, 65 female) monolingual typically developing Finnish-speaking children aged 3;6–6;7 (M = 5;0) were recruited from nurseries and pre-schools in Kotka, South-East Finland. An additional four children were excluded from the study because of experimenter error (N = 1), reluctance in taking part in testing (N = 1), and lack of concentration in the task (N = 2).

Table 1. *Overview of different target verb types*

	PT suffix	Level of productivity	No. of target verbs
Real	-si	productive	40
	-si	non-productive	8
	-i	semi-productive	40
	-i	non-productive	8
Novel	-si	productive	12
	-i	semi-productive	12
		Total	120

2.2. *Items*

One-hundred-and-twenty past tense indicative active verb forms were selected as test items (for an overview see Table 1). Each child was tested on 60 verbs (20 productive *[-si],* 20 semi-productive *[-i];* 4 non-productive -*si,* 4 non-productive -*i*; 6 novel productive -*si*, 6 novel semi-productive -*i*) of the total 120, which were generated using a Latin-square design so that each verb was presented to children as equal number of times as possible. A greater number of (semi-)productive and novel verbs were tested because it is on these verb types where the single and dual route models make different predictions (while they have similar predictions about irregular inflections). The test verbs were chosen such that they were (i) assumed to be familiar to the children, (ii) easily depicted pictorially, and (iii) represented a variable range of token frequencies and PNDs. The lemma/base frequencies of the verbs were also searched and included in the analyses. The items in the novel group were only manipulated for PND, since by definition they do not appear in any corpus. The test items were controlled for length in terms of number of phonemes (6–8 phonemes). It was impossible to keep the number of syllables constant (8/40 productive and 1/24 novel verbs taking the *i*-suffix had two syllables, the rest had three syllables). Hence, word length in syllables was included in our analysis. We discuss how we manipulated frequency and PND below.

2.2.1. *Frequency.* Type frequencies for the different verb classes were obtained from a Finnish dictionary (Suomen kielen perussanakirja 1994). Out of the 25 verb classes into which Finnish verbs can be categorised, we selected one verb class in which all verbs types take the -*si* past tense suffix only (*salata – sala+si*), and one in which all verb types take the -*i* suffix (*muistaa – muist+i*). Reflecting the overall type frequencies of -*si* and -*i*, our -*i* suffix class has distinctly more verb types (N = 2691) than the -*si* suffix class (N = 860).

The lemma frequency and past tense token frequency for each (real) verb in the *salata – sala+si* and *muistaa – muist+i* verb classes were extracted from the CSC Language Bank newspaper corpora (131.4 million word tokens, http://www.csc.fi). Based on their past tense frequency, a number of (a) productive (b) semi-productive and (c) non-productive (i.e., qualitative consonant gradation) verbs were selected. In addition to the newspaper corpora, we searched for the past tense token frequencies in a Finnish child language corpus (Kirjavainen-Max Planck corpus). The Kirjavainen-Max Planck corpus child 'Piia' was audio-recorded by her parents for approximately three separate hours a week between the ages of 1;7 and 4;0. She was the first born child in an upper working/middle class family, in Kotka, South-East Finland. The interaction during the recordings was completely spontaneous, most commonly involving a dialogue between the child and one or both of her parents during meal times or play sessions.

The Piia corpus is still in the process of being transcribed, and the morphological coding is yet to be conducted. Therefore, only the third person singular past tense token frequencies of all target verbs were searched. This was achieved using the FREQ program of the CLAN software (MacWhinney 2000), which searched for these word forms in the child and her interlocutors' data. At the time of the searches, transcribed recordings came from the period between 1;7–3;0 (approximately 277 000 word tokens). The frequency counts on Piia's data hence come from a younger child than the children that were tested in our experiment. However, as there was no other Finnish child corpus available to us, Piia's data was the best way to include frequency counts that reflect language with which a Finnish child is likely to be familiar.

2.2.2. *Phonological neighbourhood density (PND).* PND for each target verb were searched in the Finnish dictionary (Suomen kielen perussanakirja 1994). Phonological 'friends' were searched by looking for clusters of verbs that are minimal pairs, so that they differed only in the first phoneme (consonant) and rhymed with each other in the A-infinitive (e.g., *pelata, kelata, selata*) and past tense forms (e.g., *pelasi, kelasi, selasi*). Our target verbs differed in the number of phonological 'friends' but the number of enemies was kept constant as zero. As the non-productive stem change verbs have only a few or no friends we could not measure the PND on these items. The output lists of phonological friends for each test verb was examined by the first and second author, who excluded all verbs they assumed that 4–6-year-old children would not be familiar with, and hence these verbs would not contribute to the PND effect. Both the adjusted and non-adjusted numbers of phonological friends were included as variables in the PND analysis. PND ranged between 0–7 friends.

2.2.3. *Verbs types.* Based on their frequency and PND, we selected 40 productive (*-si*) and 40 semi-productive (*-i*)[8] verbs from the 860 *-si* and 2691 *-i* verbs that belonged to the relevant verb classes. Only verbs that show overt past tense marking in colloquial as well as in standard Finnish were selected as test verbs.

In addition to the productive target verbs, eight (N = 8) non-productive (i.e., qualitative consonant gradation) verbs were selected from the same *-si* category from which we selected the other *-si* suffix verbs. As there were no qualitative consonant gradation verbs in our *-i* suffix class, eight (N = 8) non-productive *-i* suffix verbs were selected from another verb class (*kuunnella – kuuntel+i*).[9] Our motivation for including non-productive verbs was twofold. First and foremost, the inclusion of them allowed us to test the effect of frequency on irregular-like items, although this was not the main focus of our investigation since both dual and single route models make the same predictions for these verbs. Second, these verb types were included to keep the children focused during testing because these items were expected to be more difficult.

To test children's productivity with the two past tense suffixes, we created 24 novel verbs (12× *-si* and 12× *-i*). This was done by changing the first phoneme (consonant) of a real (semi-)productive Finnish verb (e.g., see Niemi 2006), which allows us to predict how a given novel verb should be inflected (as the inflection of a verb is largely determined by the form of its stem). Initial bi- and trigram searches were conducted on these to ensure that the novel verbs reflected phonotactics of real Finnish words. Like real verbs, the novel verbs were manipulated for PND. The PND ranged between zero and seven friends;

8. Note that to comply with an anonymous reviewer's suggestion, we have excluded two of our semi-productive *-i* suffix verbs (*istahtaa*, *hypähtää*) from the results due to the two verbs not only showing inflectional but also derivational processes. However, we also ran a mixed effects model in which these two verbs were included in the analysis. The results of this analysis are very similar to the analysis in which they are excluded; all significant effects remain significant apart from an inhibiting surface frequency effect.
9. Note that because it was difficult to find non-productive verbs which (1) children are familiar with, (2) have suitable frequencies and (3) can be presented pictorially, many of the non-productive verbs (N = 6) were, strictly speaking, derivations. However, we are not convinced that children would treat these types of words as derivations even if an adult or a linguist would. Therefore, because our method was a non-speeded production task, because the children were given the A-infinitive form of the verb, and because the children were given two chances to produce the target verb in between which the E repeated the A-infinitive form of the target verb if needed, we do not think that the derivational properties of these verbs are problematic for our non-productive verb analysis.

none of the novel verbs had phonological enemies[10]. They came from the same two verb classes as were used in the real verb part of the study.

The test verbs are listed in Appendix B (Tables A1, A2, B and C), along with their newspaper lemma frequencies, newspaper past tense token frequencies, the token frequencies of the 3sg past tense forms in the Piia-corpus, and number of phonological friends.

2.3. *Materials and procedure*

2.3.1. *Past tense elicitation.* The children's knowledge of the past tense was investigated using a 'wug-test' elicitation task (Berko 1958). In the task, the children saw a picture on a laptop computer screen and simultaneously heard a pre-recoded test sentence describing the picture through headphones (for novel verbs, the pictures depicted novel actions) (see Appendix A for examples). Following Lyytinen (1988a, 1988b), verbs were presented in an A-infinitive form and elicited in the past tense in sentences such as (13).

(13) *Niko tykkää/haluaa pelata.*
Niko likes/wants to play.
Eilenkin/ äskenkin/ viimeksikin se _____.
Yesterday/ a moment ago/ last time he _____.

To familiarise the child with the procedure, three practice verbs were administered before the testing began: 1× *-i* suffix verb *kävellä – käveli* ('walk'), 1× *-si* or *-i* suffix verb *piirtää – piirsi/piirti* ('draw') and 1× stem change verb which could be taken as either *-si* or *-i juosta – juoksi* ('run'). All the practice items were from verb classes other than the two used in the present study and did not rhyme with any of the test verbs. The test verbs were presented in a random order selected by E-prime software (MacWhinney et al. 2001). If the child did not respond at all, produced the infinitive or present tense form, or produced a non-target verb, the experimenter emphasised that the event took place before the current time and that the child had to use the same verb as they had heard. The test sentence and the corresponding picture were then played again one more time. Each child was tested on 60 out of total of 120 items, which were counterbalanced using a Latin-square design. To avoid fatigue, testing was conducted over two sessions which were administered on different days. Verbs with similar meanings/forms (*hypätä* 'to jump' vs. *hypähtää* 'to spring') were

10. The real words that the zero-friend novel verbs derived from were such that they (i) had no friends or enemies, (ii) had not been used in the real verb part of the study, and (iii) were likely to be unfamiliar to the children (e.g., *niisittää, koulata*). These criteria meant that our novel verbs indeed had zero friends even though they rhymed with a real Finnish verb.

tested at different sessions or between different participant groups. Different pictures depicting the actions of these verbs were used.

As it is difficult to determine the phonetic quality of suffixes in child speech based on audio information alone, the children's responses were transcribed during testing when visual information about articulation was also available. In addition, the children's answers were audio recorded directly into a laptop computer using the SoundForge digital editing and recording software. The online transcriptions were checked against the recording and any discrepancies were corrected. Six percent of the children's answers (4yo: N = 4 children, 5yo: N = 3 children, 6yo: N = 1 child) were second coded to assess transcription reliability. The agreement was high: 98.5% (473/480) of the children's responses were transcribed exactly the same way. Discrepancies were adjudicated by the first author.

2.4. *Past tense coding*

Children's final responses were coded as being correct or erroneous. Responses were coded as correct when they exactly matched the standard Finnish (e.g., *pela+ta* → *pela+si*) or grammatical colloquial past tense forms (e.g., *oott+i* instead of *odott+i, pela+s* instead of *pela+si*). Any deviation from the correct form was coded as an error. Erroneous responses were further coded into the following categories:

(1) No response: the child did not say anything or did not produce a verbal response.
(2) No change (i.e., a repetition of the infinitive form of the target verb[11])
(3) Present tense responses
(4) Erroneous Past tense
    (i) wrong suffix used with the correct or near correct stem (e.g. **leim+i* for *leima+si* and **soitta+si* for *soitt+i*)
    (ii) consonant (or vowel) length errors (e.g., **pum**p**asi* for *pum***pp***asi*, **tyki**t**i* for *tyki***tt***i*)
    (iii) qualitative errors to the stem; these mostly applied to non-productive (i.e., qualitative consonant gradation) verbs.
    (iv) other past tense errors: a past tense suffix was overtly present but the form of the verb was not correct.
(5) Other

Reliabilities were computed on 5% of the data distributed evenly across ages. The agreement between the first and second coding (i.e., coding a par-

11. Note that the infinitive form of some test verbs is identical to the present tense form of those verbs. These were coded as infinitive i.e., 'no change' rather than present tense verbs.

Table 2. *Mean accuracy and standard deviation by condition*[12]

	Mean	SD
-si productive	.86	.04
-i semi-productive	.80	.07
-si non-productive	.62	.14
-i non-productive	.47	.17
-si novel	.79	.05
-i novel	.74	.08

ticular answer in the same way) was 96.73%, Cohen's Kappa = .962. Discrepancies were adjudicated by the first author.

3. Results

The overall error proportion was 22.2% (21.7% if 2-syllable verbs were excluded, see below). Table 2 presents the accuracy results for each verb class across the entire sample of children.

Table 2 shows that the children performed best overall on the most productive categories. Their performance was next best on the novel words, whereas their performance on the non-productive words was low in comparison. Children performed better overall with words taking the *-si* suffix.

We next present the data for the frequency and phonological neighbourhood density (PND) analyses for real and then the novel verbs. We then analyse the children's errors in more detail.

3.1. *Frequency and phonological neighbourhood density (PND)*

Following recent developments in the analysis of psycholinguistic data, we analysed the data using mixed effects models (see Baayen et al. 2008). The following variables served as fixed effects: (i) suffix type (*-si* vs. *-i*), (ii) log transformed lemma frequency, (iii) log transformed token frequency in the newspaper corpus, (iv) log transformed token frequency in the Piia corpus, (v) number of phonological friends (a) in a Finnish dictionary and (b) when the exhaustive list was amended for children's vocabulary, (vi) amended phonological neighbourhood density (high [$N \geq 2$ friends] vs. low [$N = 0$ friends]), (vii) syllable length of the past form, (viii) sex of the child and (ix) age (in months). We included multiple measures of frequency and PND so that our analyses would be exhaustive. Note that there was not collinearity for token

12. If the 2-syllable verbs are excluded (see description of the mixed effects models we ran) the mean and SD for semi-productive *-i* verbs is M = .82 SD = .04 and for novel *-i* verbs M = .76 SD = .04.

and lemma frequencies. Random effects were: (i) participant, (ii) item, and (iii) verb list on which the children were tested. All possible interaction terms were included. Our initial statistical analyses found that the length of verb in syllables significantly positively affected the production in all age groups. That is, all age groups made significantly more errors with two-syllable than three-syllable verbs ($p < .001$). Therefore, the eight real and one novel 2-syllable *-i* suffix verb were excluded from our frequency and PND effect analyses[13]. However, we will return to the two-syllable verbs in the error analysis and in the Discussion.

Table 3 presents the results for the variables that significantly contributed to the children's performance on the real (semi-)productive verbs.

Table 3. *Significant effects for real productive and semi-productive verbs in the overall analysis of the three age groups*

	Estimate	Std. Error	Z value	Pr(>|z|)
(Intercept)	4.03626	1.62513	2.484	0.013*
i vs. si	−0.35524	0.14880	−2.387	0.017*
Log(newspaper lemmaf + 1)	−0.32935	0.12072	−2.728	0.006*
Log(newspaper tokenf + 1)	0.32902	0.16498	1.994	0.046*
Friends (adjusted)	−0.11912	0.05167	−2.305	0.021*
Age	−0.10931	0.02523	−4.333	<0.001***

*$p < .05$, **$p < .01$, ***$p < .001$.

For the real (semi-)productive verbs we found facilitating effects of age, lemma frequency, and PND. That is, children made fewer errors with age, and fewer errors as the lemma frequency and the number of phonological friends increased. However, there was a significant inhibiting effect of token frequency in the newspaper corpus; the children performed better with low than high token frequency verbs. The children performed significantly better with words taking the *-si* suffix than the *-i* suffix. There were no interactions, suggesting that PND and lemma frequency affected all the age groups and the two suffixes (*-si* and *-i*) equally.

The significant contributing variables from the analysis of the non-productive verbs is presented in Table 4.

13. The results remain very similar even if the 2-syllable verbs were included: an analysis of real (semi-)productive *-si* and *-i* suffix verbs showed that the children performed significantly better with *-si* than *-i* verbs ($p = .003$), with high base frequency verbs ($p = .019$) and with verbs that had higher phonological neighbourhood densities ($p = .048$). Their performance in the test got better with age ($p < .001$). The production of novel verbs was affected by the children's age ($p < .001$). The children also performed significantly better with the *-si* than *-i* suffix novel verbs ($p < .001$).

Table 4. *Significant effects for non-productive verbs in the overall analysis of the three age groups*

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.73578	0.89521	4.173	<.001***
Log(newspaper lemmaf + 1)	−0.31635	0.12672	−2.496	0.013*
Age	−0.04718	0.01303	−3.62	<.001***
i vs. si	1.26809	0.92194	1.375	0.169
Age: si interaction	−0.0303	0.01398	−2.168	0.03*

*$p < .05$, **$p < .01$, ***$p < .001$.

Non-productive (i.e., qualitative consonant gradation) verbs showed a significant facilitating effect of age and lemma frequency; that is, children made fewer errors as they got older and as lemma frequency increased. There was also a significant interaction between the suffix type (*-si* vs. *-i*) and age: this reflected the fact that the rate with which the children correctly inflected verbs that take the *-si* suffix increased at a greater rate across development than their comparable performance on the *-i* suffix.

Table 5. *Significant effects for novel verbs in the overall analysis of the three age groups*

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.47845	1.26850	2.742	0.006**
Age	−0.09417	0.02108	−4.467	<.001***

*$p < .05$, **$p < .01$, ***$p < .001$.

The variables that significantly contributed to the children's performance on the novel verbs are presented in Table 5.

Table 5 shows that the children's overall performance on novel verbs was affected only by their age. That is, the children made fewer errors as they got older.

3.2. *Errors*

We now turn to our analyses of the children's erroneous responses. All items (N = 120) were included in this analysis (i.e., the two-syllable *-i* suffix verbs (N = 9) and the two derivational *-i* suffix verbs that were excluded from the mixed effects models reported above were included in the error description here). The proportion of non-past tense responses as a function of erroneous responses (N = 1830) was 53.9%. Of all responses (N = 8160), the proportion of non-past tense responses was 12.1%. Thus, it is safe to assume that overall the children understood the task.

The errors were coded according to the coding scheme outlined in the Method section. Table D in the Appendix B illustrates the error proportions of each error type for each target verb. We will next discuss each error type in turn.

(1) No response
It was relatively rare that the children did not respond at all or produced a non-verbal response (N = 24).

(2) No change
The target verb was produced in the infinitive form in which the children heard the test verb (N = 597) relatively frequently. This was regardless of the fact that after an infinitive (or a present tense) response the experimenter emphasised that the event took place prior to the moment of testing. Of the total 597, 149 of these errors were made with verbs taking the *-si* and 448 with verbs taking the *-i* suffix. An analysis of these errors using the independent variables of frequency and PND revealed no reliable effects. The difference between the suffixes can be largely explained by the fact that the infinitive used as the prompt and the third person present tense forms of the semi-productive and novel *-i* verbs are identical (e.g., *heittää*(A-inf) – *heittää*(3sg) 'to throw', while the *-si*, suffix verbs are not (e.g., *kerätä*(A-inf) – *kerää*(3sg) 'to pick'. Even though the former were coded as 'no change'-errors, it is impossible to determine whether the child was producing a repetition or a present tense form.

(3) Present tense responses
Present tense responses of the target verbs were also produced relatively frequently (N = 226). Two-hundred-and-fifteen (N = 215) of these were made with verbs taking the *-si* suffix and 11 with *-i* suffix (see the previous paragraph for an explanation for the discrepancy between the suffixes). Non-target verbs were also occasionally produced in the present tense (N = 99). These were mostly semantically congruent with the picture provided to elicit the target verb.

(4) Erroneous past tense responses
Target verbs that were produced in an erroneous past tense forms were the biggest error category (N = 636). Of these, 194 were regularization (i.e., suffix) errors and 442 were errors to the stem. Each is discussed below in detail.

(i) Suffix errors
Both *si*-for-*i* and *i*-for-*si* errors were found (N = 128, 3.1% and N = 66, 1.6% of responses overall, respectively). Sixty-two (N = 62) of the *si*-for-*i* errors occurred in the nine two-syllable words, and a *t*-test showed that the children made significantly more regularization errors with two- than three-syllable words $t(44) = 3.12$, $p < .001$. If we only include three-syllable words in the

analysis, the number/proportion of overgeneralization/irregularization errors with the two types of verbs are very similar (*si*-for-*i*: N = 66 1.9%, *i*-for-*si*: N = 66 1.6%[14]), suggesting no overall preference for children to generalize either suffix. However, similar to Dąbrowska (2008b), it seemed to be the case that some children preferred using one suffix variant, the others the other, and some children overgeneralized both. It should be noted though that not all the children produced overgeneralizations (approximately 30% of the children did when two-syllable words were excluded). Also, many of the children who did produce overgeneralization errors with the three-syllable words produced only one or two overgeneralization errors overall. Hence, making any robust claims about the regularization of one or the other or both suffixes is not possible. Nevertheless, about a third of these children only overgeneralized *i*-for-*si* contexts, a third *si*-for-*i* contexts and a third produced instances of both overgeneralization error types. Interestingly, it seemed to be the case that those children who only produced *i*-for-*si* errors were younger (mean age 54 months [4;6], SD = 9.3) than those children who only produced *si*-for-*i* errors (mean age 61 months [5;1], SD = 11.0). Children who produced both overgeneralization types slotted in between these two ages (mean age 56 months [4;8], SD = 8.6). Hence, it could be that earlier in development the children pay attention to distributional properties of suffixes and overgeneralize the most common suffix type into incorrect contexts. Later in development they may pay more attention to the segmentation of the word into syllables and morphemes (see Discussion). However, as emphasized above, the low number of overgeneralization errors produced by each individual child does not lend itself to any assertive analyses based on our data.

Finally, separate non-parametric correlations were run for the two and three syllable words. There was a significant negative correlation between regularization error rates and past tense token frequency in the newspaper corpus for three-syllable items ($r_s = -.365$, $p = .013$) and error rates and token frequency in Piia's data for two-syllable ($r_s = -.600$, $p = .044$) words. These results suggest that children are more likely to select the incorrect suffix when token frequency is low. However, the latter result is likely to be inflated because of the small sample size (N = 9).

(ii) Consonant (and vowel) length errors

Consonant (and vowel) length errors (N = 50) were infrequent (between 0–4% for most of our verb types), and occurred most commonly with -*si* suffix verbs. This can be explained by the fact that the suffixation process of sixteen of -*si* suffix target verbs result in quantitative consonant gradation (from single to

14. Because the participants were only tested on 60 items each, not every item was tested with the same frequency, leading to a slight difference in percentage error rates.

double consonant), whereas this was not the case for any of the *-i* suffix verbs. Consonant length errors were only relatively common with one real verb type (*kam**p**ata – kam**pp**asi* 'to trip someone over') (17%) and one novel verb (*kum**p**ata – kum**pp**asi*) (11%). The likely reason why the children produced the past tense form of *kampata* with a single /p/ (**kampasi*) is because this is the past tense form of a Finnish (qualitative consonant gradation) verb that was not used in our study (*ka**mm**ata – ka**mp**asi* 'to comb'). However, consonant length errors where also produced with *-i* suffix verbs, which do not exhibit consonant gradation. The direction of these was both from weak form [i.e., short consonant] to strong form [i.e., geminate consonants] (e.g., **lo**rr**o**tti* instead of *lo**r**otti* [N = 1]) and vice versa (e.g., **tyki**t**t**i* instead of *tyki**t**ti* [N = 2], **pö**l**l**ytti* instead of *pö**ll**ytti* [N = 1]).

(iii) Qualitative errors to the stem

We found one-hundred-and-thirty-two (N = 132) qualitative errors to the stem. These most commonly occurred with non-productive (i.e., qualitative consonant gradation) verbs, which the children often inflected as if they were (semi-)productive verbs (e.g., *mu**ll**a+ta – *mu**ll**a+si* instead of *mu**lt**a+si*). That is, the children treated these verbs, in which a consonant in the middle of the stem should have changed, as regular suffix-only verbs and consequently only, incorrectly, added the past tense suffix without changing the infinitive stem. This can be seen as a similar error to English-speaking children producing past tense forms like **go+ed* instead of *went*. These types of errors are well documented in Finnish children (e.g., Laalo 2011; Niemi and Niemi, 1987). Since the qualitative stem change verbs have no or at best only a couple of phonological friends, learning the inflectional process involved in these types of verbs can be expected to be difficult (unless they are of high frequency and hence rote-learned).

(iv) Other past tense errors

Two-hundred-and-sixty (N = 260) errors were coded in this category. One-hundred-and-sixteen (N = 116) of these were instances in which a past tense suffix was overtly present; however, even though the child seemed to be trying to produce the target verb, the base of the verb was not totally correct (e.g., **lasketti* or **lapsistis* instead of *lasvitti*). In addition, past tense forms of non-target verbs were also produced (N = 144). These most commonly consisted of errors in which the child produced a phonetically distinct yet semantically related verb (e.g., *otti* 'took' instead of *nappasi* 'caught'). However, with a small number of target verbs a semantically different but similar sounding verb was produced in the past tense by some children (e.g., *pelasi* 'played' instead of *pelkäsi* 'was scared of' (N = 20), *jäähdytti* 'cooled down' instead of *jäädytti* 'froze' (N = 5). For the two semi-productive verbs (*istahtaa-istaht+i* 'to sit down casually', *hypähtää-hypäht+i* 'to spring') which were excluded from the

analysis reported in Table 3, the children often exchanged the stem into a semantically and phonetically similar verb, *istui* 'sat down' (N = 8); *hyppäsi* 'jumped' (N = 6); *hyppi* 'jumped' (N = 1) respectively.

(5) Other

One-hundred-and-nineteen (N = 119) responses were coded as 'other'. These commonly included erroneous present and infinitive verb forms or other (non-past) forms of the target verbs (e.g., **luvaata* or **luvattaa* for the target verb *lupasi*). Additionally, past tense responses that did not resemble the target word and were not real Finnish verbs were coded into this category (e.g., **nasvas* or **naskas* for *lasvitti*).

3.3. *The most common error verbs*

The five most common erroneously inflected (semi-)productive verbs were all semi-productive (*-i* suffix) verbs (*liittää, istahtaa, toistaa, poistaa, hypähtää*). Overall, these were produced in an erroneous form between 29–46% of their responses. Two of these verbs (*istahtaa, hypähtää*) had very similar meaning and phonetic form to the erroneous form that the children often produced (see above). The other three were two-syllable verbs, two of which had relatively high token and lemma frequencies. These three (like the other six two syllable verbs) were often erroneously inflected using the *-si* suffix, resulting in regularization errors (e.g., *liittä+s(i)*, *toista+s(i)*) (see above). Alternatively, many children produced them in the form they had heard the verb being produced (i.e., in the infinitive).

The non-productive (i.e., qualitative consonant gradation) verbs were the most difficult verb category for the children. The error percentages for the non-productive verbs ranged between 18% and 82% (mean: 45.9%, SD: 0.178). *Madella – matel+i* (82%), *jaella – jakel+i* (66%) and *mullata – multa+si* (60%) had the highest error proportions in this category. The error type most commonly produced with non-productive verbs was stem errors in which the children inflected the verb as if it were a suffix-only verb (e.g., **mullasi* instead of *multasi*). Occasionally the children also changed the stem into a non-existing stem form (e.g., **malelli* for *mateli*) (i.e., stem form that is not the past tense form or any other real form). Thus, even though overall the children performed very well with quantitative consonant gradation verbs (e.g., *pum**p**ata → pum**pp**asi*) (see above) they found qualitative consonant gradation verbs (e.g., *mu**ll**ata → mu**lt**asi*) difficult.

The error percentages for the novel verbs ranged between 14% and 48%. The novel verb with which most errors (48%) were made was a two-syllable word *toittaa*. This *-i* suffix verb was erroneously produced with the *-si* suffix (see above). Other novel verbs were also prone to error. These included, for instance, *lasvittaa* (35%), *kumpata* (31%) and *kastata* (28%). *Kumpata* is a

quantitative consonant gradation (*-si*) verb and was typically erroneously produced with a single /p/ in the past tense (11%). *Kastata* was most commonly irregularized so that instead of the children producing the correct form *kasta+si* they produced *kastat+i* (N = 5). *Lasvitti* was often produced with changes to the stem, even though it is a semi-productive (*-i*) verb (e.g., **laskitti*, **lapsistis*, **lasvetti*) (N = 9).

4. Discussion

To test the predictions of the single and dual route models, we investigated (i) the role of frequency and phonological neighbourhood density (PND) in Finnish children's acquisition of productive, semi-productive and non-productive verbs, and (ii) the error types that the children produced. We consider how the results bear on both approaches.

4.1. *Dual and single route model predictions*

Finnish past tense production does not provide much support for the dual route model's predictions (e.g., Marcus et al. 1995; Pinker 1991; Pinker and Prince 1998; Pinker and Ullman 2002). First, the fact that we found a PND effect for the (semi-)productive (*-si* and *-i*) verbs is not predicted by the dual route model, since such effects are indicative of processing in associative memory. The single route approach, on the other hand, predicts that these words should show PND effects. This adds to previous studies that have found PND effects in less inflected languages (Kidd and Lum 2008; Marchman 1997; Ragnarsdóttir et al. 1999).

Second, while the lemma frequency was facilitating, suggesting that the children composed (rather then memorised) past tense word forms, we found a significant inhibiting effect of token frequency. That is, the children tended to perform better with low than high frequency (semi-)productive verbs. Clahsen et al. (2004) also observed this 'anti-frequency' effect in a speeded production experiment with children and adults, whereby their participants produced low frequency items quicker than high frequency ones. They interpreted the effect to be consistent with the dual route model, whereby the effect is attributed to competition between the lexical access and computational routes. More specifically, they suggest that with high frequency regulars both the lexical and computational routes get activated. The activation of the stored (lexical) item turns off the computational route and hence slows down the computation of those words. This type of a competition, which assumes that the speaker's language processing system is divided between retrieving the stored item and the computation of a complex form, may have resulted in interference and consequently brought about the anti-frequency effect in our non-speeded production task. Note, however, that this effect is also predicted by single route

approaches, which explicitly allow for both full-form storage and schema application. The difference between the two explanations is epistemological: the dual-route model assumes competition between storage and abstract rule-driven processing, whereas the single route approach assumes competition between storage and composition via schema-based or analogical processing.

We also found the lemma frequency to affect the processing of non-productive verbs. This finding is difficult to explain by the dual route model, since the effect of lemma frequency suggests that the children engaged in a suffixation process with irregular-type verbs. On the other hand, the single route model, which makes no clear distinction between regular and irregular words but which allows for composition to take place (e.g., Smolka et al. 2007), can account for these results.

Further problematic issues for the dual route account come from our over-generalization/irregularization data. Some dual route proponents assume that there is only one default suffix (e.g., Ullman 2001), whereas some acknowledge that there might be more than one (e.g., Pinker 1999; Pinker and Prince 1994). Hence, the dual route accounts predict that (a) the Finnish children make either *si*-for-*i* or *i*-for-*si* errors (Ullman 2001) or (b) the Finnish children may make both (Pinker 1999; Pinker and Prince 1994). We found that the children produced both error types and that there was no significant difference in the *si*-for-*i* and *i*-for-*si* error rates. By the same token, the dual route models predict that the novel items should be inflected by (a) using one or the other of the past tense suffixes, or (b) by both. We found that both the -*si* and -*i* suffixes were productively used with novel verbs at levels above chance. These results suggest that there are two default past tense processes in Finnish. However, the fact that we found a PND effect for both productive (-*si*) and semi-productive (-*i*) verbs means that under the dual route model neither process can be the default. As the single route model does not assume a separate default process but instead that all verb types are processed by the same associative system, these results are more consistent with the single route approach.

4.2. *Type frequency*

The children performed better on the -*si* than -*i* suffix real (semi-)productive verbs. Note that the two syllable (-*i* suffix) words with which the children made distinctly more errors were removed from these analyses and therefore the difficulty associated with two syllable past tense -*i* verbs was not causing this effect. Hence, it seems that the children in our study, like children in some previous studies (e.g., Niemi 2006), find the production of past tense items that take the -*si* suffix easier than those that take the -*i* suffix. What can explain this difference? The dual route account might suggest that the -*si* suffix is the default and therefore the children perform better with it. However, this is not a

reasonable explanation given that (i) both past tense suffixes in Finnish are productive, (ii) they both showed PND effects (iii) both were correctly used with novel verbs above chance, and (iv) both were overgeneralized into incorrect contexts. That is, we found no indication of different processes being involved for the two verb types. Hence, the dual route model struggles to explain the better performance with *-si* than *-i*.

On the other hand, the single route model assumes that type frequency has the single strongest effect on acquisition in terms of productivity (e.g., Bybee 1995 and see support for this suggestion e.g., Albright 2002). Hence, if the *-si* suffix has a higher type frequency than the *-i* suffix, the single route account would expect that the *-si* suffix is generalized more readily than the *-i* suffix. However, the *-si* suffix verb class that we used in our study had a distinctly lower number of verb types than our *-i* suffix verb class (N = 860 vs. 2691). This reflects the type frequency of the two suffixes in Finnish overall (N = 1406 vs. 7479). Therefore, type frequency cannot explain the difference in performance between *-si* and *-i* suffixes (for similar findings in German noun plurals see e.g., Laaha et al. 2006; Marcus et al. 1995). On the other hand, it has been suggested that morphemes that are more easily segmented from the word are more easily identified and acquired (e.g., Ragnarsdóttir et al. 1999; Slobin 1973). Niemi (2006) explained his Finnish participants' better performance on *-si* than *-i* suffix verbs with the easier identification process of the *-si* suffix. We suggest this is also the case with the verb classes used in our study. The morpheme and syllable boundary in the *-si* suffix past tense verbs is the same, making the *-si* suffix relatively easy to segment from the word. The *-i* suffix, on the other hand, is not a free-standing syllable, but needs to be incorporated into the last sound(s) of the stem (e.g., *soit-t+i*) to make the final syllable. This incorporation is likely to be a more difficult process, making the segmentation of this suffix more difficult (e.g., Dressler 1985). The fact that we found both *si*-for-*i* and *i*-for-*si* errors suggests that the higher type frequency of the *-i* suffix contributes to its productivity regardless of its more difficult segmentation. Nevertheless, type frequency alone does not determine which Finnish past tense suffix/verb class is the most productive.

4.3. *Two-syllable words*

We found that significantly more errors were produced with two- than three-syllable verbs even though the majority of the two-syllable verbs had high token frequencies. The overall error percentage for the two-syllable words responses was 29.2% (N = 183/627). This finding, where the length of the target word had a facilitating effect, is somewhat surprising. In erroneous responses to two-syllable verbs the children commonly produced either a repetition of the A-infinitive form of the verb they heard as the prompt (51.9%,

N = 95/183) or a *si*-for-*i* error (33.9%, N = 62/183). As all of the two-syllable verbs that we used take the *-i* suffix, the only overgeneralization that can be made with them is in this direction. The erroneous *-si* suffixation in the two-syllable words commonly means that a two-syllable word (e.g., *soittaa* 'to play') becomes a three-syllable word (*soit-ti* vs. **soit-ta-si*). One reason for these errors may be priming of three-syllable verbs during testing, as the majority (N = 111) of the test verbs had three syllables. If the errors are caused by this, the priming must be at a relatively abstract three-syllable level, as the three-syllable target verbs represented a number of word types at vowel (V) and consonant (C) level (e.g., CV-CVC-CV, CVC-CVC-CV, VC-CVC-CV). To investigate whether priming during testing could have caused these errors, we coded for the number two-syllable word *si*-for-*i* errors in the first and last half of each testing block. We found that the *si*-for-*i* errors with two-syllable words were produced roughly equally in the two testing blocks (N = 28 first; 34 last), suggesting that activation of a three-syllable representation during testing does not account for these errors.

Alternatively the children may have a strong three syllable schema for past tense forms. Such an explanation would fit with a schema-based model if three syllable words were the most common type of past form in the ambient language. To test this suggestion, we counted the number of syllables in past tense form of the 500 most frequent verbs-lemmas in the Finnish news paper corpus (CSC Language Bank newspaper corpora). The results are presented in Table 6. We found that three syllable structure was the most common (54%) past tense form, but that two syllable past tense verbs were also relatively common (37%). However, the fact that we coded for the highest frequency verbs rather than low frequency or a random sample of verbs may have given an overestimation of the proportion of two-syllable verbs, since high frequency words are often shorter than lower frequency items (Bybee 1988, 2006). It is therefore safe to assume that the most common past tense verb forms have three-syllables, which may be one reason why the children transformed two-syllable

Table 6. *The number and proportion of different syllable structures of the past tense forms of the 500 highest frequency finnish verb lemmas*

number of syllables	N	%
5-syllables	2	<1%
4-syllables	28	6%
3-syllables	271	54%
2-syllables	183	37%
1-syllable	12	2%
excluded	4	1%
Total	500	100%

words into three-syllable words. However, it is important to point out that, in colloquial speech, one can often omit the final /i/ during *-si* suffixation. This is the case for adult-adult language (Hakulinen et al. 2005: 180) as well as Child Direct Speech. A rough frequency count in the CDS in Kirjavainen-Max Planck Finnish child corpus on the production of *-si* vs. *-s* realization of 20 of our high frequency *-si* suffix test verbs showed that the colloquial verb forms (i.e., *-s* instead of *-si*) was used 91% of the time. Our participants showed a similar preference for using the colloquial suffix form (61%, N = 38/62). Crucially, this means that the majority of the children's *si*-for-*i* errors with two-syllable words resulted in an incorrect *two-*, not three-, syllable word because the *-s* suffix does not stand alone as a syllable but is instead incorporated in the last syllable of the word (e.g., **soit-tas* instead of *soit-ta-si*). Therefore, even though the type frequency of three syllable words in Finnish in general may be a contributing factor to the overgeneralization errors with two-syllable words, it is not just a simple case of children preferring three syllables over two.

Lastly, one factor contributing to the overgeneralization of the *-si* suffix into *-i* context with two-syllable words may be caused by incorrect analogy. Examples (14)–(16) show that the 3sg present tense forms of our two-syllable *-i* (e.g., *soit-taa*) and three-syllable *-si* verbs (e.g., *pump-paa*) have two syllables, while the three-syllable *-i* suffix verbs (e.g., *vil-kut-taa*) have three syllables.

	A-INF	**3sg PRES**	**3sg PAST**	**GLOSS**
(14)	*pum-pa-ta*	***pump-paa***	***pump-pa-si***	to pump
(15)	*soit-taa*	***soit-taa***	****soit-ta-si***	to play
(16)	*vil-kut-taa*	*vil-kut-taa*	*vil-kut-t+i*	to wave

Hence, at the syllable level the two-syllable *-i* verbs bear similarities to the three-syllable *-si* verbs in the 3sg present tense but differ from the three syllable *-i* suffix verbs from their own verb class. This similarity with the *-si* suffix verbs (and lack of similarity to the *-i* suffix verbs) may have caused the errors. This suggestion is supported by Räisänen's (1975) observations on his twin sons' language development (see also e.g., Karlsson 1983 where he argues that 3sg indicative active [present tense] is the source form for verb inflections). Räisänen suggested that children may use any inflected form that they have acquired to incorporate the lexical item into an existing paradigm—and not necessarily the A-infinitive form we used as the prompt. The children who participated in our study may have considered the similarity between the target verb and other verbs in their vocabulary also in the 3sg present tense forms when determining verb class membership for the target verbs and their consequent inflection. It is also worth bearing in mind that, in terms of their level of openness (i.e., productivity), the *-si* suffix class(es) are more open for new

words than the *-i* suffix class(es). The combination of the level of productivity of the *-si* suffix and the two syllable verbs' similarity to verbs taking the *-si* suffix is likely to have contributed to the children treating the two-syllable *muist+aa – muist+i* verbs as the three-syllable *sala+ta – sala+si* verbs (for similar observations with other *-i* suffix classes see Niemi 2006).

5. Conclusion

In the current study we tested the competing predictions of the single- and dual-route explanations of past tense acquisition in Finnish. For productive and semi-productive verbs we found lemma frequency and phonological neighbourhood density effects. For non-productive (i.e., qualitative consonant gradation) verbs only lemma frequency effects were found. We also found similar overgeneralization/irregularization error rates, and productivity with novel items for both past tense suffixes. These data best support the predictions of the single route approach. In contrast, the results pose a number of problems for the application of the dual route account to Finnish.

Received 16 July 2010

Accepted 20 November 2011

University of Manchester, UK
University of Eastern Finland
The Australian National University

Appendices

Appendix A

Example Pictures of Real Verbs

Kahlata (‘to paddle’)

Kuunnella (‘to listen’)

Example Pictures of Novel Verbs

Vököttää (to meek)

Novauttaa (‘to gorp’)

Appendix B. Frequency and phonological neighbourhood density statistics for the items used in the study

Note: Frequency estimates were taken from (i) the CSC Language Bank newspaper corpora (131.4 million word tokens, www.csc.fi/), and (ii) the Kirjavainen-Max Planck corpus of one child collected longitudinally between ages 1;7–4;0. At the time of the searches the corpus constituted 277 000 word tokens. PND estimates were taken from a Finnish dictionary (Suomen kielen perussanakirja 1994) and were adjusted by the first and second author by removing any neighbours that children aged 4–6-years were unlikely to know.

Table A1. *Productive -si suffix verbs, their lemma and token frequencies and phonological neighbourhood densities (adjusted and non-adjusted)*

INF	PT form	Gloss	Lemma freq/ million	Log Lemma freq/million + 1	PT token freq/ million	Log PT freq/ million + 1	PT token freq in the Piia corpus/million	Non-adjusted PND	Adjusted PND
pelata	pelasi	to play	472.38	6.16	70.99	4.28	14.44	5	3
herätä	heräsi	to wake up	76.45	4.35	9.95	2.39	79.42	2	2
vastata	vastasi	to answer	479.34	6.17	55.65	4.04	0	2	2
leikata	leikkasi	to cut	100.6	4.62	3.74	1.56	21.66	3	3
viitata	viittasi	to put hand up/refer to	101	4.62	16.64	2.87	0	3	2
kerätä	keräsi	to collect	206.57	5.34	34.69	3.57	0	2	2
haukata	haukkasi	to bite	3	1.39	0.46	0.38	18.05	2	2
halata	halasi	to hug	3.59	1.52	0.46	0.38	14.44	2	2
pilata	pilasi	to ruin	26.57	3.32	4.61	1.72	0	3	2
hukata	hukkasi	to lose	13.48	2.67	3.32	1.46	0	3	2
pakata	pakkasi	to pack	24.35	3.23	1.58	0.95	3.61	6	4
selata	selasi	to flick through	4.37	1.68	0.24	0.22	0	5	3
jumpata	jumppasi	to exercise	1.95	1.08	0.04	0.04	0	4	4
pumpata	pumppasi	to pump	8.96	2.3	0.4	0.34	0	4	4
kipata	kippasi	to tip	2.32	1.2	0.21	0.19	0	3	2
keilata	keilasi	to bowl	5.06	1.8	1.88	1.06	0	4	3

tankata	tankkasi	to fill up	5.49	1.87	0.31	0.27	0	3	2
kampata	kamppasi	to trip over	0.35	0.3	0.06	0.06	0	3	2
tiskata	tiskasi	to wash up	1.57	0.94	0.04	0.04	0	2	2
surrata	surrasi	to buzz	0.78	0.58	0.08	0.08	0	3	3
lisätä	lisäsi	to add	441.11	6.09	32.07	3.5	0	0	0
törmätä	törmäsi	to crash into	54.82	4.02	18.5	2.97	0	0	0
ohjata	ohjasi	to steer	188.1	5.24	14.53	2.74	0	0	0
korjata	korjasi	to fix	120.76	4.8	7.88	2.18	7.22	0	0
napata	nappasi	to catch	35.75	3.6	16.19	2.84	14.44	1	0
maalata	maalasi	to paint	51.16	3.95	5.81	1.92	18.05	2	0
arvata	arvasi	to guess	36.83	3.63	0.91	0.65	7.22	0	0
kurkata	kurkkasi	to peek	0.66	0.51	0.02	0.02	10.83	1	0
hypätä	hyppäsi	to jump	55.34	4.03	13.34	2.66	64.98	1	0
kiusata	kiusasi	to tease	24.92	3.26	2.15	1.15	0	0	0
leimata	leimasi	to stamp	36.56	3.63	4.26	1.66	0	0	0
murskata	murskasi	to crush	14.33	2.73	2.21	1.17	0	0	0
paikata	paikkasi	to patch	18.66	2.98	1.72	1	0	0	0
hölkätä	hölkkäsi	to jog	1.61	0.96	0.21	0.19	0	0	0
tuhlata	tuhlasi	to spend	10.87	2.47	0.74	0.55	0	0	0
kahlata	kahlasi	to paddle	3.9	1.59	0.34	0.29	0	0	0
kärrätä	kärräsi	to push (wheel barrow)	2.33	1.2	0.18	0.17	0	0	0
filmata	filmasi	to film	4.22	1.65	0.28	0.25	0	0	0
vispata	vispasi	to whisk	1.16	0.77	0.26	0.23	0	0	0
putsata	putsasi	to clean	4.44	1.69	0.4	0.34	0	0	0

Table A2. *Semi-productive -i suffix verbs, their lemma and token frequencies and phonological neighbourhood densities (adjusted and non-adjusted)*

INF	PT form	Gloss	Lemma freq/ million	Log Lemma freq/million + 1	PT token freq/million	Log PT freq/ million + 1	PT token freq in the Piia corpus/million	Non-adjusted PND	Adjusted PND
voittaa	voitti	to win	478.36	6.17	216.85	5.38	21.66	2	2
näyttää	näytti	to show	567.1	6.34	77.34	4.36	0	2	2
heittää	heitti	to throw	131.59	4.89	31.47	3.48	86.64	2	2
soittaa	soitti	to ring/play	192.77	5.27	27.61	3.35	104.69	2	2
toistaa	toisti	to repeat	36.63	3.63	12.38	2.59	7.22	3	2
pysäyttää	pysäytti	to stop	49.31	3.92	8.54	2.26	0	2	2
poistaa	poisti	to erase	119.15	4.79	6.24	1.98	7.22	3	2
tiputtaa	tiputti	to drop	4.51	1.71	1.01	0.7	36.1	2	2
tupsahtaa	tupsahti	to appear	3.33	1.47	0.97	0.68	14.44	4	4
vilkuttaa	vilkutti	to wave	2.56	1.27	0.58	0.46	14.44	3	2
liittää	liitti	to attach	63.99	4.17	3.22	1.44	0	4	3
täräyttää	täräytti	to bang	5.05	1.8	3.11	1.41	0	4	4
mököttää	mökötti	to sulk	0.39	0.33	0.02	0.02	0	6	6
käkättää	käkätti	to cackle	0.19	0.17	0.02	0.02	0	4	3
kilkuttaa	kilkutti	to tinkle (e.g., piano)	0.16	0.15	0.02	0.02	0	3	2
rimputtaa	rimputti	to ring/play	0.05	0.05	0.02	0.02	0	3	2
hytkyttää	hytkytti	to shake	0.05	0.05	0.01	0.01	0	4	3
köröttää	körötti	to trundle along	0.14	0.13	0.01	0.01	0	3	3
lorottaa	lorotti	to run liquid	0.39	0.33	0.01	0.01	0	3	2
vollottaa	vollotti	to howl	0.15	0.14	0.01	0.01	0	4	2
esittää	esitti	to act	673.63	6.51	89.01	4.5	21.66	0	0
yllättää	yllätti	to surprise	158.23	5.07	28.72	3.39	14.44	0	0
herättää	herätti	to wake someone up	132.19	4.89	26.26	3.31	14.44	0	0
odottaa	odotti	to wait	566.62	6.34	16.46	2.86	18.05	0	0
syöttää	syötti	to feed	27.01	3.33	10.31	2.43	7.22	2	0
lämmittää	lämmitti	to heat up	28.93	3.4	3.36	1.47	14.44	0	0

pamahtaa	pamahti	to bang	2.64	1.29	1.25	0.81	25.27	0	0
istahtaa	istahti	to sit down	6.03	1.95	1.06	0.72	14.44	1	0
värittää	väritti	to colour	8.33	2.23	0.85	0.62	14.44	0	0
hypähtää	hypähti	to spring	1.51	0.92	0.59	0.46	14.44	0	0
kannustaa	kannusti	to cheer	49.55	3.92	4.6	1.72	0	0	0
peruuttaa	peruutti	to reverse	25.44	3.27	4.57	1.72	0	0	0
selostaa	selosti	to commentate	16.45	2.86	4.56	1.72	0	0	0
venyttää	venytti	to stretch	11.5	2.53	2.97	1.38	0	0	0
tykittää	tykitti	to bomb	4.69	1.74	2.61	1.28	0	0	0
jäädyttää	jäädytti	to freeze	14.24	2.72	2.24	1.18	0	0	0
virnistää	virnisti	to grin	4.51	1.71	0.9	0.64	0	0	0
kalastaa	kalasti	to fish	15.6	2.81	0.68	0.52	0	0	0
pöllyttää	pöllytti	to dust	1.48	0.91	0.65	0.5	0	0	0
lävistää	lävisti	to pierce	2.41	1.23	0.58	0.46	0	0	0

Table B. *Non-productive (i.e., qualitative consonant gradation) verbs, their lemma and token frequencies and phonological neighbourhood densities (adjusted and non-adjusted)*

INF form	PT form	Gloss	PT suffix	Lemma freq/million	PT token freq/million	PT token freq in Piia/million	Non-adjusted PND	Adjusted PND
annella	anteli	to give	i	0.08	0.00	0.00	0	0
huudella	huuteli	to shout	i	5.21	0.72	14.44	1	1
jaella	jakeli	to hand out	i	5.94	1.31	0.00	0	0
kierrellä	kierteli	to wander/circle	i	16.24	2.72	0.00	0	0
kuunnella	kuunteli	to listen	i	107.67	6.60	0.00	1	0
madella	mateli	to crawl	i	1.98	0.73	0.00	2	2
työnnellä	työnteli	to push	i	0.79	0.12	0.00	1	0
vaihdella	vaihteli	to exchange	i	98.32	5.52	0.00	0	0
jahdata	jahtasi	to chase	si	11.70	0.78	0.00	3	2
karata	karkasi	to run away	si	42.73	16.84	0.00	3	0
levätä	lepäsi	to rest	si	26.38	1.49	0.00	0	0
liata	likasi	to dirt/soil	si	2.91	0.09	0.00	0	0
luvata	lupasi	to promise	si	231.60	52.31	0.00	1	0
maata	makasi	to lie down	si	22.62	2.71	0.00	3	2
mullata	multasi	to soil	si	0.93	0.00	0.00	3	1
pelätä	pelkäsi	to be scared of	si	173.00	7.52	0.00	0	0

Table C. *Novel verbs and their phonological neighbourhood densities (adjusted and non-adjusted)*

INF form	PT form	PT suffix	Lemma freq/million	PT token freq/million	Non-adjusted PND	Adjusted PND
keinittää	keinitti	i	0	0	0	0
kiisittää	kiisitti	i	0	0	0	0
kolvittaa	kolvitti	i	0	0	0	0
lasvittaa	lasvitti	i	0	0	0	0
märäyttää	märäytti	i	0	0	5	5
näkistää	näkisti	i	0	0	0	0
novauttaa	novautti	i	0	0	0	0
pöröttää	pörötti	i	0	0	4	4
pupsahtaa	pupsahti	i	0	0	5	5
sytkyttää	sytkytti	i	0	0	5	4
toittaa	toitti	i	0	0	3	3
vökköttää	vökötti	i	0	0	7	7
hyölätä	hyöläsi	si	0	0	0	0
kalskata	kalskasi	si	0	0	0	0
kastata	kastasi	si	0	0	3	3
kennata	kennasi	si	0	0	0	0
kumpata	kumppasi	si	0	0	5	5
liskata	liskasi	si	0	0	3	3
netlata	netlasi	si	0	0	0	0
riehätä	riehäsi	si	0	0	0	0
toulata	toulasi	si	0	0	0	0
turrata	turrasi	si	0	0	4	3
veilata	veilasi	si	0	0	5	4
velata	velasi	si	0	0	6	4

Table D. *The error proportions of each error type and for each target verb*

Verb	PT suffix	No response	No change	PRES responses	Erroneous PT responses				Other	TOTAL
					Wrong suffix	C/V length	Stem error	Other PT errors		
Productive										
arvata	si	0.01	0.07	0.01	0.01	0.00	0.00	0.01	0.00	0.13
filmata	si	0.00	0.03	0.06	0.02	0.00	0.00	0.03	0.02	0.15
halata	si	0.00	0.02	0.08	0.00	0.00	0.00	0.02	0.00	0.11
haukata	si	0.00	0.00	0.09	0.00	0.00	0.00	0.03	0.00	0.12
herätä	si	0.00	0.03	0.09	0.00	0.00	0.00	0.00	0.00	0.12
hölkätä	si	0.00	0.04	0.04	0.04	0.04	0.00	0.01	0.03	0.21
hukata	si	0.00	0.04	0.03	0.00	0.00	0.00	0.01	0.03	0.11
hypätä	si	0.00	0.02	0.11	0.00	0.02	0.02	0.02	0.00	0.17
jumpata	si	0.00	0.00	0.10	0.00	0.01	0.00	0.00	0.00	0.11
kahlata	si	0.02	0.05	0.06	0.00	0.00	0.00	0.00	0.00	0.12
kampata	si	0.00	0.02	0.08	0.00	0.17	0.00	0.00	0.00	0.26
kärrätä	si	0.00	0.05	0.08	0.05	0.00	0.00	0.02	0.02	0.20
keilata	si	0.00	0.05	0.06	0.00	0.00	0.00	0.02	0.00	0.12
kerätä	si	0.00	0.03	0.06	0.00	0.00	0.00	0.00	0.00	0.09
kipata	si	0.00	0.04	0.04	0.01	0.04	0.00	0.00	0.00	0.14
kiusata	si	0.00	0.03	0.06	0.05	0.00	0.00	0.00	0.00	0.14
korjata	si	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.01	0.08
kurkata	si	0.00	0.03	0.09	0.02	0.03	0.00	0.00	0.02	0.18
leikata	si	0.00	0.03	0.04	0.00	0.01	0.01	0.00	0.00	0.10
leimata	si	0.00	0.04	0.06	0.01	0.00	0.00	0.01	0.01	0.14
lisätä	si	0.00	0.03	0.08	0.00	0.00	0.00	0.03	0.02	0.15
maalata	si	0.00	0.01	0.08	0.03	0.00	0.00	0.01	0.01	0.15
murskata	si	0.00	0.03	0.06	0.06	0.00	0.00	0.07	0.01	0.23
napata	si	0.00	0.06	0.06	0.02	0.02	0.02	0.05	0.03	0.25

ohjata	si	0.00	0.01	0.06	0.01	0.00	0.00	0.03	0.01	0.13
paikata	si	0.00	0.06	0.06	0.01	0.01	0.00	0.03	0.00	0.17
pakata	si	0.00	0.01	0.03	0.01	0.01	0.00	0.00	0.01	0.08
pelata	si	0.00	0.03	0.06	0.00	0.00	0.00	0.01	0.01	0.11
pilata	si	0.00	0.03	0.11	0.00	0.00	0.00	0.00	0.00	0.14
pumpata	si	0.00	0.02	0.09	0.00	0.02	0.00	0.00	0.00	0.12
putsata	si	0.00	0.03	0.06	0.02	0.00	0.00	0.03	0.00	0.14
selata	si	0.01	0.08	0.03	0.00	0.00	0.00	0.01	0.00	0.14
surrata	si	0.00	0.02	0.08	0.03	0.00	0.00	0.02	0.02	0.15
tankata	si	0.00	0.03	0.10	0.00	0.00	0.00	0.00	0.00	0.13
tiskata	si	0.00	0.02	0.06	0.00	0.00	0.00	0.05	0.00	0.12
törmätä	si	0.00	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.06
tuhlata	si	0.01	0.03	0.04	0.03	0.00	0.00	0.00	0.00	0.11
vastata	si	0.01	0.08	0.04	0.00	0.00	0.00	0.04	0.01	0.20
viitata	si	0.00	0.06	0.00	0.04	0.01	0.00	0.03	0.03	0.17
vispata	si	0.00	0.00	0.12	0.00	0.02	0.00	0.05	0.02	0.20
Semi-productive										
esittää	i	0.00	0.13	0.01	0.00	0.00	0.00	0.00	0.00	0.14
heittää	i	0.00	0.11	0.00	0.07	0.00	0.01	0.00	0.03	0.23
herättää	i	0.00	0.10	0.00	0.00	0.00	0.00	0.01	0.01	0.13
hypähtää	i	0.00	0.06	0.08	0.00	0.00	0.00	0.16	0.00	0.29
hytkyttää	i	0.01	0.11	0.01	0.00	0.00	0.00	0.03	0.01	0.18
istahti	i	0.00	0.11	0.05	0.00	0.00	0.00	0.21	0.02	0.38
jäädyttää	i	0.00	0.12	0.03	0.00	0.00	0.00	0.08	0.00	0.23
käkättää	i	0.00	0.15	0.00	0.00	0.00	0.00	0.02	0.02	0.18
kalastaa	i	0.00	0.08	0.00	0.07	0.00	0.00	0.03	0.00	0.18
kannustaa	i	0.01	0.11	0.00	0.01	0.00	0.00	0.00	0.01	0.15
kilkuttaa	i	0.02	0.09	0.02	0.00	0.00	0.02	0.05	0.00	0.18
köröttää	i	0.00	0.10	0.00	0.00	0.00	0.00	0.04	0.03	0.17
lämmittää	i	0.00	0.15	0.00	0.00	0.00	0.00	0.02	0.00	0.17

Table D. (*Continued*)

Verb	PT suffix	No response	No change	PRES responses	Erroneous PT responses				Other	TOTAL
					Wrong suffix	C/V length	Stem error	Other PT errors		
lävistää	i	0.02	0.05	0.06	0.03	0.00	0.00	0.05	0.02	0.22
liitti	i	0.03	0.14	0.03	0.21	0.00	0.01	0.01	0.03	0.46
lorottaa	i	0.00	0.09	0.03	0.00	0.02	0.00	0.08	0.02	0.23
mököttää	i	0.00	0.10	0.03	0.00	0.00	0.00	0.04	0.01	0.18
näyttää	i	0.00	0.14	0.00	0.05	0.00	0.00	0.00	0.00	0.18
odottaa	i	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.01	0.11
pamahtaa	i	0.00	0.15	0.02	0.00	0.00	0.00	0.00	0.00	0.17
peruuttaa	i	0.00	0.08	0.00	0.00	0.01	0.00	0.01	0.01	0.13
poisti	i	0.02	0.20	0.02	0.05	0.00	0.00	0.03	0.00	0.31
pöllyttää	i	0.00	0.10	0.00	0.01	0.01	0.00	0.00	0.03	0.15
pysäyttää	i	0.00	0.14	0.02	0.00	0.02	0.02	0.00	0.00	0.18
rimputtaa	i	0.00	0.11	0.03	0.00	0.00	0.00	0.03	0.00	0.17
selosti	i	0.00	0.14	0.01	0.04	0.00	0.00	0.03	0.04	0.27
soittaa	i	0.00	0.13	0.00	0.01	0.00	0.01	0.01	0.00	0.17
syöttää	i	0.00	0.13	0.00	0.03	0.00	0.00	0.03	0.01	0.20
täräyttää	i	0.00	0.08	0.01	0.00	0.00	0.00	0.04	0.01	0.15
tiputtaa	i	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.14
toisti	i	0.00	0.20	0.00	0.15	0.00	0.00	0.01	0.01	0.38
tupsahtaa	i	0.00	0.14	0.01	0.01	0.00	0.00	0.07	0.01	0.25
tykittää	i	0.00	0.15	0.00	0.02	0.03	0.00	0.02	0.00	0.22
värittää	i	0.00	0.09	0.00	0.00	0.00	0.00	0.02	0.00	0.11
venyttää	i	0.00	0.15	0.00	0.00	0.00	0.03	0.03	0.00	0.22
vilkuttaa	i	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.17
virnistää	i	0.02	0.09	0.00	0.00	0.00	0.03	0.05	0.02	0.20

voittaa	i	0.00	0.18	0.00	0.03	0.00	0.00	0.00	0.00	0.21
vollottaa	i	0.00	0.11	0.05	0.00	0.02	0.00	0.00	0.00	0.17
yllättää	i	0.00	0.15	0.00	0.01	0.00	0.00	0.03	0.00	0.20
Non-productive										
jahdata	si	0.00	0.00	0.12	0.05	0.00	0.00	0.02	0.00	0.18
karata	si	0.00	0.06	0.06	0.00	0.00	0.10	0.03	0.03	0.27
levätä	si	0.00	0.02	0.12	0.02	0.00	0.14	0.02	0.02	0.32
liata	si	0.02	0.02	0.08	0.03	0.00	0.28	0.03	0.02	0.46
luvata	si	0.01	0.03	0.01	0.00	0.00	0.11	0.03	0.06	0.25
maata	si	0.00	0.17	0.07	0.00	0.01	0.23	0.06	0.01	0.55
mullata	si	0.00	0.05	0.05	0.02	0.00	0.46	0.02	0.02	0.60
pelätä	si	0.01	0.04	0.03	0.00	0.00	0.04	0.30	0.01	0.44
annella	i	0.02	0.06	0.12	0.08	0.00	0.25	0.03	0.00	0.55
huudella	i	0.00	0.04	0.04	0.07	0.00	0.37	0.06	0.00	0.58
jaella	i	0.00	0.02	0.09	0.14	0.00	0.37	0.03	0.02	0.66
kierrellä	i	0.01	0.07	0.00	0.01	0.00	0.35	0.04	0.04	0.54
kuunnella	i	0.00	0.04	0.03	0.01	0.00	0.01	0.07	0.01	0.18
madella	i	0.05	0.06	0.06	0.20	0.00	0.32	0.02	0.11	0.82
työnnellä	i	0.00	0.02	0.11	0.03	0.00	0.32	0.00	0.00	0.48
vaihdella	i	0.01	0.01	0.06	0.01	0.00	0.34	0.01	0.01	0.46
Novel										
hyölätä	si	0.00	0.07	0.03	0.03	0.01	0.00	0.06	0.04	0.24
kalskata	si	0.00	0.03	0.03	0.11	0.00	0.00	0.04	0.04	0.25
kastata	si	0.00	0.02	0.08	0.08	0.00	0.00	0.08	0.03	0.28
kennata	si	0.00	0.02	0.06	0.05	0.00	0.00	0.02	0.03	0.17
kumpata	si	0.00	0.05	0.06	0.02	0.11	0.00	0.06	0.02	0.31
liskata	si	0.00	0.07	0.03	0.04	0.00	0.00	0.06	0.03	0.23
netlata	si	0.00	0.03	0.11	0.02	0.02	0.00	0.02	0.03	0.22
riehätä	si	0.00	0.03	0.06	0.02	0.00	0.00	0.05	0.05	0.20
toulata	si	0.00	0.06	0.03	0.00	0.00	0.00	0.04	0.04	0.17

Table D. (*Continued*)

Verb	PT suffix	No response	No change	PRES responses	Erroneous PT responses				Other	TOTAL
					Wrong suffix	C/V length	Stem error	Other PT errors		
turrata	si	0.00	0.07	0.03	0.00	0.00	0.00	0.04	0.00	0.14
veilata	si	0.00	0.03	0.07	0.01	0.00	0.00	0.06	0.00	0.17
velata	si	0.00	0.05	0.02	0.03	0.02	0.00	0.06	0.05	0.22
keinittää	i	0.00	0.08	0.00	0.08	0.00	0.00	0.01	0.03	0.21
kiisittää	i	0.00	0.12	0.00	0.02	0.00	0.00	0.08	0.03	0.25
kolvittaa	i	0.00	0.15	0.02	0.05	0.00	0.00	0.03	0.02	0.26
lasvittaa	i	0.01	0.06	0.01	0.00	0.00	0.00	0.17	0.10	0.35
märäyttää	i	0.00	0.07	0.01	0.01	0.00	0.00	0.06	0.04	0.20
näkistää	i	0.00	0.17	0.02	0.00	0.00	0.00	0.02	0.03	0.23
novauttaa	i	0.00	0.15	0.02	0.02	0.02	0.00	0.03	0.00	0.23
pöröttää	i	0.00	0.12	0.02	0.00	0.03	0.00	0.05	0.00	0.22
pupsahtaa	i	0.01	0.10	0.01	0.01	0.00	0.00	0.06	0.03	0.23
sytkyttää	i	0.00	0.11	0.00	0.02	0.02	0.00	0.06	0.05	0.25
toittaa	i	0.00	0.14	0.01	0.28	0.00	0.00	0.03	0.01	0.48
vököttää	i	0.00	0.11	0.03	0.00	0.00	0.00	0.06	0.00	0.20
Mean		0.00	0.07	0.04	0.02	0.01	0.03	0.03	0.01	0.22

References

Albright, Adam. 2002. Islands of reliability from regular morphology: Evidence from Italian. *Language* 78. 684–709.

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90. 119–161.

Baayen, R. Harald, Doug J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59. 390–412.

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

Bybee, Joan L. 1988. Morphology as lexical organization. In Michael Hammond & Michael Noonan (eds.), *Theoretical morphology: Approaches in Modern Linguistics*, 119–141. London: Academic Press.

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

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82. 711–733.

Bybee, Joan & Jean Newman. 1995. Are stem changes as natural as affixes? *Linguistics* 33. 633–654.

Bybee, Joan L. & Slobin, Dan I. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58. 265–289.

Cathey, James E. & Dreidre Wheeler. 1986. Finnish verbal morphophonology and consonant gradation. *Nordic Journal of Linguistics* 9. 103–133.

Clahsen, Harald, Meike Hadler & Helga Weyerts. 2004. Speeded production of inflected words in children and adults. *Journal of Child Language* 31. 683–712.

Dąbrowska, Ewa. 2001. Learning a morphological system with a default: The Polish genitive. *Journal of Child Language* 28. 545–574.

Dąbrowska, Ewa. 2004. Rules or schemas? Evidence from Polish. *Language and Cognitive Processes* 19. 225–271.

Dąbrowska, Ewa. 2008a. The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language* 58. 931–951.

Dąbrowska, Ewa. 2008b. The later development of an early-emerging system: The curious case of the Polish genitive. *Linguistics* 46. 629–650.

Dressler, Wolfgang U. 1985. On the predictiveness of Natural Morphology. *Journal of Linguistics* 21. 321–337.

Eddington, David. 2009. Spanish verbal inflection: a single- or dual-route system? *Linguistics* 47. 173–199.

Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho. 2005. *Iso Suomen Kielioppi* [The large grammar of Finnish]. Helsinki: Suomalaisen Kirjallisuuden Seura.

Hartshorne, Joshua K. & Michael T. Ullman. 2006. Why girls say 'holded' more than boys. *Developmental Science* 9. 21–32.

Jaggar, Philip J. 2001. *Hausa*. Amsterdam, Philadelphia: John Benjamins.

Karlsson, Fred. 1977. Eräistä morphologian teorian ajankohtaisista ongelmista [On some current problems in morphological theory]. *Sananjalka* 19. 26–56.

Karlsson, Fred. 1983. *Suomenkielen äänne- ja muotorakenne* [Structure of Finnish. Phonology and Morphology]. Helsinki: WSOY.

Karlsson, Fred & Kimmo Koskenniemi. 1985. A process model of morphology and lexicon. *Folia Linguistica* XIX(1–2). 217–231.

Kidd, Evan & Jarrad A. G. Lum. 2008. Sex differences in past tense overregularization. *Developmental Science* 11. 882–889.

Laaha, Sabine, Dorid Ravid, Katharina Korecky-Kröll, Gregor Laaha & Wolfgang U. Dressler. 2006. Early noun plurals in German: Regularity, productivity or default? *Journal of Child Language* 33. 271–302.

Laalo, Klaus. 2011. *Lapsen varhaiskielioppi ja minipardigmat.* [Child's early grammar and miniature-paradigms]. Helsinki: Suomalaisen Kirjallisuuden Seura.

Langacker, Roland W. 1987. *Foundations of Cognitive Grammar: Vol. 1. Theoretical Prerequisites.* Stanford: Stanford University Press.

Langacker, Roland W. 2000. A dynamic usage-based model. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based Models of Language*, 1–63. Stanford: CSLI Publications.

Lyytinen, Paula. 1988a. Suomen kielen taivutusmuotojen hallinta 2–9-vuotiailla [Finnish 2–9-year-olds' proficiency with inflections]. *Reports from the Department of Psychology University of Jyväskylä* 297.

Lyytinen, Paula. 1988b. Morfologiatesti [Morphology test]. *Reports from the Department of Psychology University of Jyväskylä* 298.

MacWhinney, Brian. 2000. *The CHILDES project.* Hillsdale: Erlbaum.

MacWhinney, Brian, James S. James, Chris Schunn, Ping Li, & Walter Schneider. 2001. STEP—A system for teaching experimental psychology using E-Prime. *Behavior Research Methods, Instruments and Computers* 33. 287–296.

Marchman, Virginia A. 1997. Children's productivity in the English past tense: The role of frequency, phonology, and neigborhood structures. *Cognitive Science* 21. 283–304.

Marchman, Virginia A. & Elizabeth Bates. 1994. Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language* 21. 339–366.

Marcus, Gary, Ursula Brinkmann, Harald Clahsen, Richard Wiese & Steven Pinker. 1995. German inflection: The exception that proves the rule. *Cognitive Psychology* 29. 189–256.

Marcus, Gary F., Steven Pinker, Michael T. Ullman, Michelle Hollander, T. John Rosen & Fei Xu. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57 (4, Serial No. 228): 1–165.

Martin, Maisa. 1989. *Amerikansuomen Morfologiaa ja Fonologiaa* [approximate translation: The Morphology and Phonology of the American-Finnish]. Licentiate thesis. Department of Finnish language, University of Jyväskylä, Finland.

Niemi, Jussi. 2006. A paradigm competition: an experimental note on Finnish verbs. In Mikael Suominen, Antti Arppe, Anu Airola, Orvokki Heinämäki, Matti Miestamo, Urho Määttä, Jussi Niemi, Kari K. Pitkänen & Kaius Sinnemäki (eds.), *A man of Measure: Festschrift in Honour of Fred Karlsson on his 60th Birthday. Special Supplement to SKY Journal of Linguistics* 19. 227–235.

Niemi, Jussi & Sinikka Niemi. 1987. Acquisition of Inflectional Marking: A case study of Finnish. *Nordic Journal of Linguistics* 10. 59–89.

Pinker, Steven. 1991. Rules of language. *Science* 253 (5019). 530–535.

Pinker, Steven. 1999. *Words and Rules*. London: Phoenix.

Pinker, Steven & Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28. 73–193.

Pinker, Steven & Alan Prince. 1994. Regular and irregular morphology and the psychological status of rules of grammar. In Susan D. Lima, Roberta Corrigan and Gregory K. Iverson (eds.), *The Reality of Linguistic Rules*. 321–351. Amsterdam: John Benjamins.

Pinker, Steven & Michael T. Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6. 456–463.

Plunkett, Kim & Virginia Marchman. 1993. From rote learning to system building. Acquiring verb morphology in children and connectionist nets. *Cognition* 48. 21–69.

Plunkett, Kim & Ramin C. Nakisa. 1997. A connectionist model of the Arabic plural system. *Language and Cognitive Processes* 12. 807–838.

Ragnarsdóttir, Hrafnhildur, Hanne G. Simonsen & Kim Plunkett. 1999. The acquisition of past tense morphology in Icelandic and Norwegian children: An experimental study. *Journal of Child Language* 26. 577–618.

Ravid, Dorid, Wolfgan U. Dressler, Bracha Nir-Sagiv, Katharine Korecky-Kröll, Agnita Souman, Katja Rehfeldt, Sabine Laaha, Johannes Bertl, Hans Basbøll and Steven Gillis. 2008. Core morphology in child directed speech; crosslinguistic corpus analyses of noun plurals. In Heike Behrens (ed.), *Corpora in Language Acquisition Research*. Amsterdam: John Benjamins.

Rumelhart, David E. & James L. McClelland. 1986. On learning the past tenses of English verbs. In David E. Rumelhart & James L. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2. Psychological and Biological Models*. Cambridge: MIT Press.

Räisänen, Alpo. 1975. Havaintoja lastenkielestä [Observations on child language]. *Virittäjä* 79. 251–266.

Slobin, Dan I. 1973. Cognitive prerequisites for the development of grammar. In Charles A. Ferguson and Dan I. Slobin (eds.), *Studies of Child Language Development*, 175–208. New York: Holt, Rinehart & Winston.

Smolka, Eva, Pienie Zwisterlood & Frank Rösler. 2007. Stem access in regular and irregular inflection: Evidence from German participles. *Journal of Memory and Language* 57. 325–347.

Storkel, Holly L. 2004. Do children acquire dense neighbourhoods? An investigation of similarity neighbourhoods in lexical acquisition. *Applied Psycholinguistics* 25. 201–221.

Suomen Kielen Perussanakirja [Finnish Language Dictionary]. 1994. Helsinki: Edita Oyj.

Ullman, Michael T. 1999. Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighbourhood effects. *Language and Cognitive Processes* 14. 47–67.

Ullman, Michael T. 2001. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30. 37–69.

Ullman, Michael T. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92. 231–270.

Xu, Fei & Steven Pinker. 1995. Weird past tense forms. *Journal of Child Language* 22. 531–556.