L2 English derivational knowledge: Which affixes are learners more likely to recognise?[1]

Dmitri Leontjev

University of Jyväskylä, Finland
dmitri.leontjev@jyu.fi

Abstract

Knowledge of derivational morphology is considered an important aspect of vocabulary knowledge both in L1 (mother tongue) and L2 (second or foreign language) English language learning. However, it is still not clear whether different derivational affixes vary in their (learning) difficulty. The present study examines whether Bauer and Nation's (1993) teaching order of L2 English affixes can account for the difficulty learners have with recognising the affixes. The participants in the study were L1 Estonian and Russian learners of English at upper-secondary schools in Estonia (*n* = 62). Their performance was measured on a word segmentation task. There were significant differences in the number of affixes the learners were able to successfully recognise at different levels, as classified by Bauer and Nation (1993). By and large, with the exception of no significant difference between Level 5 and Level 6 affixes, the higher the affix level was, the less likely the learners were to recognise the affixes at this level. I argue that these results can support the order proposed by Bauer and Nation. The implications of the finding for teaching and further research are also discussed.

Keywords: derivational morphology; affix difficulty; L2 English teaching

---

[1] The present article was used as a part of Leontjev (2016), the author's compilation dissertation.

1. Introduction

A number of studies have revealed that L2 (second or foreign language) inflectional morphology poses problems for learners (e.g., Clahsen, Felser, Neubauer, Sato, & Silva, 2010; Felser & Clahsen, 2009; Jiang, 2004; Lardiere, 1998). At the same time, while L2 learners (and native speakers alike) face even greater problems with derivational morphology (Friedline, 2011; Schmitt & Meara, 1997; Schmitt & Zimmermann, 2002; Silva & Clahsen, 2008), not many studies on learners' word derivation knowledge and its acquisition have been conducted.

Friedline (2011, p. 60) suggests that the reason for the small number of studies on word derivation has been, until recently, the predominance of theories that argue for a clear dichotomy in morphology, such as split morphology hypothesis (Perlmutter, 1988), which states that whereas inflection is rule-based, derivation only occurs in the lexicon. Nevertheless, more recent advances in morphology research, especially in the field of psycholinguistics (e.g., Alegre & Gordon, 1999; Clahsen & Neubauer, 2010), suggest that at least some derived words can be processed within the same rule-based system as (some) inflected words. Therefore, more research into word derivation and its acquisition is necessary, the more so as many questions, including how exactly learners acquire L2 word derivation knowledge, remain unanswered.

As regards L2 inflection, some relatively early morphological studies (e.g., Bailey, Madden, & Krashen, 1974), but also later studies (e.g., Pienemann, 1998), sought an answer to the question of whether there is a universal order of acquisition of L2 inflectional morphemes. With a similar objective in mind, using the research findings on the English affixes available at that time, Bauer and Nation (1993) classified L2 English affixes (both derivational and inflectional) into seven levels. The levels ranged from considering each form a different word (Level 1), to classical roots and affixes (Level 7). Later, Nation (2001) refined the classification, adding a number of affixes to the levels and limiting the list to derivational affixes only.

Bauer and Nation (1993) suggested that the levels could be used as a framework for teaching/learning affixes for reading in English. They further proposed that the levels could reflect what should be included in word families at different levels of learners' morphological awareness and be used as a reference point in empirical research on the development of word derivation knowledge. Nevertheless, up until the present time, this order has not been unambiguously confirmed or rejected empirically both as a difficulty order and the order in which learners do indeed acquire derivational affixes, or at least some of their aspects.

The present study endeavours to find evidence for Bauer and Nation's (1993) proposal, aiming at confirming that the levels they defined reflect the increasing

difficulty learners have with recognising the affixes. Therefore, the research question the study aims to answer is:

- Does the difficulty learners have with recognising derivational affixes differ significantly across the affix levels as classified by Bauer and Nation (1993), increasing as the level grows?

I will discuss Bauer and Nation's (1993) study in some detail and other research relevant for the present study in the following section. I will then present the study and the analyses, report on the findings, and suggest some research to follow which could reinforce the findings.

2. Background

In the present section, I will provide further details on Bauer and Nation's (1993) and Nation's (2001) teaching order of L2 English affixes as well as discuss the studies that used their classification or tried to challenge it. I will also discuss some (further) factors that can offer an explanation for the difficulty learners have with word derivation. Hereinafter in the paper, the levels will be referred to as Bauer and Nation's levels.

2.1. Bauer and Nation's affix levels

Bauer and Nation (1993) based their classification of affixes on the following criteria:

- frequency,
- productivity,
- predictability of the meaning of the affix,
- regularity of the written/spoken form of the base,
- regularity of the spelling/phonological form of the affix,
- regularity of function.

Table 1 Difficulty order of L2 English affixes (Bauer & Nation, 1993; Nation, 2001)

Level no.	Description
Level 1	A different form is a different word.
Level 2	Regularly inflected words are part of the same family, e.g., *-ed, -ing, -s*, etc.
Level 3	The most frequent and regular derivational affixes: -able, *-er, -ish, -less, -ly, -ness, -th* (*fourth*), *-y, non-, un-* (*unusual*)*
Level 4	Frequent and regular affixes, e.g., *-al* (*coastal*), *-ation, -ful, -ism, -ist, -ity, -ise* (*-ize*), *-ment, -ous, in-**
Level 5	Infrequent but regular affixes, e.g., *-age, -al* (*arrival*), *-ance, -ant, -ship, en-, mis-, un-* (*untie*), etc.
Level 6	Frequent but irregular affixes, e.g., *-ee, -ic, -ify, -ion, -ition, pre-, re-*, etc.
Level 7	Classical roots and affixes, e.g., *-ate, -ure*, etc.

Note. *All with restricted uses; see Appendix 1 in Bauer and Nation (1993) for details.

It is evident, and the authors themselves acknowledged it, that the criteria are not unique to Bauer and Nation's (1993) study. Similar criteria were found

to explain the acquisitional order of inflectional affixes (e.g., Goldschneider & DeKeyser, 2001) but were also used much earlier, for example, by Thorndike (1942). As the levels were defined with recognition/understanding during reading in mind, the priority was given to the written forms. The levels as identified by Bauer and Nation (1993) are presented in Table 1.

Both Bauer and Nation (1993) and Nation (2001) stressed that there was no empirical evidence for the order. On the other hand, the authors encouraged researchers to use the levels as a reference for affix difficulty in their studies. I will discuss the studies that utilised Bauer and Nation's levels in the following subsection.

2.2. Studies using Bauer and Nation's levels

Bauer and Nation's levels have been used to operationalise L2 English affix difficulty in several studies. Schmitt and Meara (1997), for example, used the levels when creating their instruments in a longitudinal study of 95 learners of English. Being the first to test the interplay between different aspects of vocabulary knowledge empirically, the authors used word derivation knowledge as one of these aspects. There was a significant, albeit small, increase in the participants' suffix knowledge over the course of the academic year (5% in the productive measure and 4% in the receptive one). The authors, however, did not find any noticeable differences between the suffixes in terms of their difficulty, which could have resulted from the fact that they used only two or three different suffixes at each level and only one Level-7 suffix.

Similarly, Schmitt and Zimmermann (2002) used the levels to control for the difficulty of the word forms across the word classes in their instrument. However, the aim of their study was to find out which parts of speech learners were the most likely to produce. Thus, the authors neither presented any data that could allow for making assumptions regarding the difficulty their participants had with affixes at Bauer and Nation's different levels nor discussed their data in terms of a potential implicational order of derivational affixes.

The authors also considered possible reasons for the difficulty that word derivation poses to L2 learners. They drew on the work of Jiang (2000), according to whom syntactic and especially morphological specifications are integrated into the lexical entry during the last stage of learning a word. Drawing on the morpheme acquisition studies (e.g., Larsen-Freeman, 1976), natural language acquisition studies (e.g., Lardiere, 1998), and psycholinguistic research (e.g., Gollan, Foster, & Frost, 1997), Jiang (2000) also claimed that by the time this latter stage is reached, many words have become fossilised.

This was an important point raised. As a matter of fact, there is psycholinguistic research demonstrating that, at least in oral-aural processing, L2 learners

often process meaning before they process form and often rely on lexical and semantic cues rather than morphological and syntactic cues during lexical processing (e.g., VanPatten, 1996). Jiang's lexical development model can serve as an explanation for that finding. On the other hand, there is also research (e.g., Clahsen & Neubauer, 2010) showing that L2 learners rely on frequency when processing derived words, that is, they process more frequent words as wholes, which can explain the usual superiority of processing meaning over processing form discussed by VanPatten (1996) and expands on Jiang's (2000) model. Specifically, more frequent L2 words may be stored and processed as wholes (perhaps due to fossilisation, according to Jiang), whereas attempts are made to analyse less frequent words, which also presents a difficulty for learners in the light of Jiang's discussion. Clahsen and Neubauer's (2010) finding is also in line with the declarative-procedural model (e.g., Ullman, 2004), according to which there are two systems involved in processing: procedural, which is rule-based and includes the processing of both inflection and derivation, and declarative for storing/retrieving frequent lexical entries as wholes. These studies present a rather strong case for controlling for frequency in word derivation research, suggesting that the more frequent morphologically complex L2 words are, the less likely they are to be analysed by learners.

Another study that used Bauer and Nation's levels was conducted by Hayashi and Murphy (2010). Their study aimed at comparing the ability to derive words of L1 (mother tongue) Japanese learners of English (*n* = 22) and adult native speakers of English (*n* = 20). The study also aimed at finding a relation between learners' size of vocabulary and their morphological awareness. The authors used affixes from Bauer and Nation's different levels as a way to establish the frequency and productivity of the affixes they used in the instruments: a word segmentation task as a measure of receptive morphological awareness and an affix elicitation task as a productive measure of it.

The authors did not elaborate on their decision to use a word segmentation task as a measure of receptive morphological awareness. However, Friedline (2011), for example, used a similar format in one of his instruments, asking the participants to write the base forms of the given derived words. He drew, above all, on the findings of Carlisle (2000) and Carlisle and Fleming (2003), which confirmed the prediction made in Schreuder and Baayen's (1995) model of morphological processing that children are able to define novel morphologically complex words in their mother tongue when they have access to corresponding bases and bound morphemes. Despite the lack of research confirming whether the same is true for L2 English, word segmentation/decomposition task types seem to be useful for establishing whether L2 learners have access to / can recognise affixes and bases in English morphologically complex words.

Hayashi and Murphy (2010) also considered semantic transparency of the items, that is, the degree to which the meaning of a whole morphologically complex word can be understood from the meaning of its parts, as one of the factors. They checked whether semantic transparency influenced their participants' morphological awareness, as previous research (e.g., Marslen-Wilson, 2007) has demonstrated that semantic transparency, among other factors, influences the processing of morphologically complex words. Having completed the qualitative evaluation of the participants' performance, Hayashi and Murphy (2010) discovered that semantic transparency influenced the way the learners performed on the word segmentation task. Specifically, they found that all of their participants were able to correctly separate affixes in the items *disorder* (Level 7), *enable* (Level 5), *rewrite* (Level 6) and *childhood* (Level 5), which were formed with the help of affixes at Bauer and Nation's different levels but which, arguably, were all semantically transparent. Judging by this finding, but also by the previous studies, semantic transparency should be taken into consideration in word derivation research, especially if it aims at establishing an implicational order of derivational affixes.

To my knowledge, there are two studies that attempted to find a difficulty order (or an order of acquisition) of L2 English derivational affixes. One of them was the study of 403 Japanese learners of English conducted by Mochizuki and Aizawa (2000). The authors evaluated the learners' knowledge of suffixes and prefixes on two nonword tasks, operationalising suffix knowledge as the ability to identify the part of speech formed with the help of the suffixes and prefix knowledge as the receptive knowledge of the meaning of the prefix. They also had an interesting way of defining affix acquisition, suggesting that affixes known by more learners are acquired earlier whereas those known by fewer learners, later. The affix order they established had several discrepancies with Bauer and Nation's levels, for example, suffix -*er* (Level 3) being more difficult that suffix -*ation* (Level 4).

One of the issues that Mochizuki and Aizawa's (2000) study had was the authors' operationalisation of suffix knowledge. It is logical to assume that syntactic function of affixes should be a part of learners' word derivation knowledge. However, limiting word derivation knowledge to syntactic function only seems to be an overgeneralisation. Moreover, the way the authors defined affix acquisition should rather be considered the order of difficulty the learners had with the affixes. Finally, as the authors mentioned, the order they established could have been affected by English loan words in Japanese.

The second study that aimed at finding a difficulty order of derivational suffixes (among other research questions) was Chuenjundaeng's (2006) master's thesis. For this purpose, the researcher used an instrument consisting of

two translation tasks including 16 base and 16 derived forms (eight base and eight derived forms per task), the base form in one task being the derived form in the other and vice versa. The second task was set a week after the first one. The suffixes that the author selected for the instrument were *-er*, *-tion*, *-ment*, and *-ity*. The Thai learners of English (*n* = 167) were asked to provide a definition/translation of the words in the tasks in their mother tongue. Their responses to each item were classified into four categories: 1 when they provided definitions for both the base and the derived form, 2 when they provided definition for only the base form, 3 when it was only the derived form, and 4 when they failed to define both the base and the derived form. The responses in Category 1 were awarded the score of 2, that is, 1 point for both the base and the derived form each. The responses in Categories 2 and 3 were awarded the score of 1, that is, 1 point for either the base or the derived form. The score of 0 was given for the items in the last category. The author used the composite score on Categories 1 and 3 as an indication of the learners' knowledge of the derived words formed with the four affixes she studied. Based on the results, the author identified the following increasing difficulty order of the suffixes: *-tion* (the total score on Categories 1 and 3 being 216), *-er* (the total score of 206), *-ity* (the total score of 154), and *-ment* (the total score of 143).

There are discrepancies between the order found by Chuenjundaeng (2006) and that found by Mochizuki and Aizawa (2000). On the other hand, it is hard to say whether the difficulty order found by Chuenjundaeng (2006) agrees with Bauer and Nation's (1993) levels or not, as all the suffixes that the author selected except for *-er* were at Level 4 of Bauer and Nation's classification. Moreover, the difference between the scores on *-er* and *-tion* was small, and the author treated *-ation*, *-ion*, and *-ition* as allomorphs of the same suffix at Level 4 of Bauer and Nation's classification. The latter is not entirely incorrect. Indeed, Bauer and Nation themselves discussed the issue of the suffix *-ation* and its allomorphy and admitted the problem of determining whether *-ation*, *-ion*, and *-ition* should be considered the allomorphs of the same suffix or not. They, however, decided that only *-ation* should be included at Level 4.

Given the lack of evidence for (or against) the order proposed by Bauer and Nation (1993), the present study sets to determine whether Bauer and Nation's levels reflect the increasing difficulty learners have with L2 English derivational affixes. However, unlike the two studies discussed above, instead of studying learners' performance on separate affixes, I will consider the affixes at each level as a group.

3. Methodology

3.1 Materials

To answer the research question, I analysed the learners' performance on a word segmentation task, a task type also used by Hayashi and Murphy (2010) and somewhat similar to Friedline's (2011) decomposition task. The purpose of the task was to find out how likely learners were to recognise affixes at Bauer and Nation's different levels. However, instead of trying to challenge Bauer and Nation's classification by studying separate affixes, as the previous studies did, I looked at the affixes at each of Bauer and Nation's level collectively, studying affixes at each of the levels as a group.

Another difference from the previous studies, specifically Friedline (2011) and Hayashi and Murphy (2010), concerned the items used in the instrument and the learners' selection procedure. What neither study did was control for the possibility that the items might have been known by their participants and thus exhibited the frequency effect. Moreover, Friedline (2011) did not control for the potential effect of the semantic transparency. Hayashi and Murphy (2010) did account for that. However, they did not establish whether the words they selected as the items in their instruments were semantically transparent to their participants and instead rated the items themselves. As with any judgmental phenomenon, not only could dissimilar ratings have been produced by other raters, but it was also not known whether the learners were actually able to discern the meanings of the words that the authors rated as semantically transparent from the meanings of the bases and the affixes these words were composed from. That is to say, it is not known whether these words were semantically transparent to the participants in their study.

I addressed the issue differently, and instead of producing figures for the frequency and the semantic transparency of the items and controlling for these while analysing the learners' performance, I made sure that the learners did not know the words selected for the task before I started the analyses.

For the item selection, I used Affix Levels @ Frequency Tester instrument from Compleat Lexical Tutor website (http://www.lextutor.ca/cgi-bin/morpho/fam_affix/index.pl). This instrument classifies words, or, rather, base words and word families in the British National Corpus into frequency bands by thousand most frequent words/word families. The instrument contains the first twenty thousand most frequent word families, breaking them into 20 frequency bands. It then separately lists derived words formed with affixes at Bauer and Nation's different levels at each of the frequency bands. That is to say, it allows for singling

out words of a certain frequency (or, rather, frequency of their bases) formed with affixes at Bauer and Nation's particular levels.

For the word segmentation task, I decided to select words formed with affixes at Bauer and Nation's Level 3 to Level 6. A total of 12 words per affix level were randomly selected such that there were three words formed with affixes at each of Bauer and Nation's levels selected at each of 5,000, 6,000, 7,000, and 8,000 most frequent base words/word families. There were, however, several constraints to the otherwise random selection.

First of all, there were not more than three words formed with the same affix, and before the selection started, words that might have been known to the participants were removed. Thus, for example, words denoting languages, such as *Croatian*, were excluded. Secondly, words containing two suffixes, such as *momentousness*, were excluded as well. At the same time, the instrument included four words formed with both a prefix and a suffix.

A further six words not including any derivational affixes were selected to serve as distractors. All in all, the instrument included a total of 50 items, of which 44 were formed with the help of a total of 48 affixes, of 10 of them being prefixes. With the exception of Level 5, there were two prefixes per Bauer and Nation's level. Since there are considerably more prefixes at Level 5 than at the rest of Bauer and Nation's levels, having four items formed with prefixes at Level 5 reflected this overall tendency. After the selection, the order of the items was randomised. The items in the present version of the instrument are presented in Table 2.

To make sure that the learners did not know any of the words in the task, and thus to account for a possible frequency effect, the participants were asked to supply translations or definitions for the items. Arguably, this also allowed for the control of whether the items were semantically transparent for the learners, on the assumption that they would supply a definition or translation for any item the meaning of which they could deduce from the meaning of the bases and the affixes.

Identifying bases could have compromised the results as this could have allowed the learners to separate affixes without actually having to recognise them. Therefore, a decision was made to exclude the performance of those learners whose translations or definitions indicated that they identified any of the bases.

Asking the participants to find prefixes and suffixes invariably meant that they had to refer to their metalinguistic knowledge to complete the task, and metalinguistic knowledge has been found to present a problem even to native speakers (e.g., Alderson, Clapham, & Steel, 1997). On the other hand, the advantages of the format, above all, the possibility to control for the influence of the participants' vocabulary knowledge, and the suitability of the task type for answering the research question, that is, finding out how well learners are able to recognise derivational affixes at Bauer and Nation's different levels, outweighed this limitation in my opinion.

Table 2 Items in the word segmentation task

Level	Frequency bands				
	5,000	6,000	7,000	8,000	9,000/10,000
Level 3	indiscreetly	blandness	briskness	brimless	
	unambiguous*	crofter	decipherable	stoutly	
	voidable	obscenely	lushness	unshackle	
Level 4	boastful	enshrinement	arsonist	discernment	
	indiscreetly	inapt	frugality	pailful	
	unambiguous*	reaffirmation	solemnise	slanderous	
Level 5	discipleship	enshrinement	deference	bestowal	
	interlace	errant	enmesh	deflationary	
	moisten	misapprehend	repentant	ternary	
Level 6**	exemption	digression	eviction	detainee	regressive
	heraldic	prohibitive	mortify		cherubic
	obstructive	reaffirmation	recoup		
Distractors	abolish, bulletin, comprise, magnitude, mediocre, scrutiny				

Notes. *Unambiguous (with a bound base morpheme) replaced an item that was found unacceptable during the piloting; **two items from the 9,000 and 10,000 most frequent word families and the item *mortify* (with a bound base morpheme) were added to replace three items that were found unacceptable during the piloting.

The instrument was piloted among five learners of English whose proficiency on the Common European Framework of Reference (CEFR; Council of Europe, 2001) was at about B1 level, as they studied at Grades 9 to 11 of Estonian schools (see *Põhikooli riiklik õppekava õigusakt*, 2010). The reason for selecting learners at this particular level of proficiency for the piloting (and also for the present study) was Nation's (2001) suggestion that the best time for starting teaching L2 affixes to learners would be when they are at the lower-intermediate level of their L2 proficiency, that is, at about level B1 on the CEFR scale.

A major aim of the piloting was to establish whether learners would be able to recognise the affixes without being able to define the words or their bases. Based on the results of the piloting, several items were replaced as one or several learners provided translations or definitions for these words. Among the items that had to be replaced were three words at Level 6. The problem I faced at this stage was that at the base frequency bands selected for creating the instrument, there were no alternatives to the selected words as the rest either contained the suffix *-ion*, which three other items in the instrument already contained, or were easy to define (e.g., *atomic* or *combative*). Thus I decided to select two items at Frequency bands 9 and 10 instead (i.e., among the 9,000 and 10,000 most frequent word families) and selected *mortify* as a new item at Frequency band 7, which is not listed in the Affix Levels @ Frequency Tester instrument but nevertheless belongs to this frequency band (the Affix Levels @ Frequency Tester instrument does not include words containing bound base morphemes).

For a similar reason, *unambiguous* replaced the item *unconstitutional*. The items used in the present version of the word segmentation task are presented in Table 2 and the Appendix.

3.2 Participants

The participants in the study were 76 L1 Estonian and L1 Russian learners of English studying at Grade 10 in Estonian schools. However, 14 of them supplied more or less accurate translations or definitions to one or several items or their bases (e.g., *wetness* for the item *moisten*), so their results were not included in the analysis. The final sample included a total of 62 learners from six different groups taught by five different teachers.

Although the proficiency level of most participants was expected to be at about level B1 on the CEFR scale, this assumption was corroborated by asking the learners to self-assess their writing and reading proficiency using the CEFR descriptors from the self-assessment scale (available at http://www.keelemapp.ee/keelemapp/keelemapi-osad/). Furthermore, the teachers were also asked to assess their learners' reading and writing proficiency using the same scale. Then, the median across the four ratings was calculated and used as the measure of the learners' proficiency in the study. While such judgmental figures should not be considered very reliable since it was just a background variable, it was found sufficient for the purposes of the present study. Moreover, the agreement between the ratings of the learners and the teachers as calculated by Kendall's tau b was substantial: $r_k = .603$, $p < .001$ for reading and $r_k = .492$, $p < .001$ for writing, which added to the reliability of the figures. Further details of the participants are presented in Table 3.

Table 3 Description of the participants

L1	N	CEFR level*				Median CEFR level
		A2	B1	B2	C1	
Estonian	27	3	17	6	1	B1
Russian	35	12	14	9	-	B1

Note. *The proficiency estimate was calculated as the median across the learners' self-assessment of their reading and writing proficiency on the CEFR self-assessment scale and the assessment of their reading and writing proficiency on the same scale completed by their teachers.

3.3. Procedure

Before the start of the study, the learners were informed that the study aimed at finding out how well they could recognise suffixes and prefixes in English. They were then given a written description of the study, which also detailed that they

were expected to assess their abilities with the help of a self-assessment scale and complete one exercise. The learners were also informed that in the study, group results rather than those of individual learners would be analysed. Then they gave their permission to use their performance for calculating the group statistics.

In order for the learners not to get discouraged by failing to define all (or, at least, most) of the words in the task, they were instructed that they should not be worried if they did not know any of the words in the task, as the task was rather difficult. The same was restated in the written instructions (see the Appendix).

It was also stressed that the learners were expected to work individually and that there was no purpose in cheating. In addition, the teachers were asked to help monitor the learners' performance. The task was not speeded, and, as the piloting had also established, it took the learners about 20 minutes to complete.

4. Results

In the present section, the results of the study will be presented. First, I will present the overall results, including the reliability estimate for the task. Following that, I will present the results that allowed me to find the answer to the research question posed in the study.

As has been mentioned earlier, the performance of 62 learners was analysed. The learners' performance was scored such that each correctly recognised affix was awarded 1 point.

Cronbach's alpha for the 48 items was .89, which suggested that the internal consistency of the instrument was rather high. Moreover, none of the learners scored 0, and there was no item in the task in which none of the learners was able to recognise the affix. This showed that there was no floor effect observed in the task. The mean number of affixes recognised by the learners was 21, $SD = 8.2$, with the weighted average of 34.85 in the 95th percentile, which means that those who scored the highest on the task were able to recognise affixes in about 73% of the items, that is, there was no ceiling effect either.

The L2 English proficiency of the two L1 groups of learners was roughly the same (see Section 3.2). Moreover, the L1 Estonian learners ($n = 27$) did not perform significantly differently from the L1 Russian learners ($n = 35$) on the segmentation task either, as demonstrated by an independent-samples t test, $t(60) = 0.61$, $p = .55$. Thus, in the following, for the most part, the performance of the two groups will be considered together. I will, however, corroborate the main analysis by comparing the two L1 groups, too.

To discover whether the learners were able to recognise affixes at Bauer and Nation's different levels to a different degree, a composite score was calculated separately for the affixes at each of Bauer and Nation's levels, the maximum

possible score being 12 at each level. The descriptive statistics are presented in Table 4. For the sake of comparison, I also supplied the means and the medians separately for each of the L1 groups.

Table 4 Affixes correctly recognised at Bauer and Nation's (1993) different levels ($n = 62$; $k = 12$ at each of the levels)

Level no.	Mean		95 % CI of the mean		SD	Median
			Lower	Upper		
Level 3	Estonian ($n = 27$)	9.48				11
	Russian ($n = 35$)	8.49				9
	Both L1s	8.92	8.08	9.76	3.29	10
Level 4	Estonian ($n = 27$)	6.04				7
	Russian ($n = 35$)	5.31				5
	Both L1s	5.63	5.03	6.23	2.37	5
Level 5	Estonian ($n = 27$)	3.41				3
	Russian ($n = 35$)	3.09				3
	Both L1s	3.23	2.68	3.77	2.16	3
Level 6	Estonian ($n = 27$)	2.78				2
	Russian ($n = 35$)	3.54				2
	Both L1s	3.21	2.54	3.88	2.66	2

From the descriptive statistics, it can be deduced that with the exception of almost no difference between Level 5 and Level 6 affixes, the numbers of affixes recognised at Bauer and Nation's different levels were rather different, and the higher the level was, the fewer affixes were recognised. Specifically, on average, the learners recognised about 75% of all the affixes at Level 3, about a half at Level 4, and about a quarter at Levels 5 and 6.

However, from the descriptive statistics, it was not clear whether the differences between the levels were statistically significant. Thus, a repeated measures ANOVA was conducted with the number of affixes recognised at Bauer and Nation's different levels forming the within-subjects factor. The ANOVA, with the Greenhouse-Geisser correction of the degrees of freedom applied as the sphericity assumption was violated, confirmed that there was a significant difference in the learners' ability to recognise derivational affixes at Bauer and Nation's different levels, $F(2.44, 149.08) = 117.66$, $p < .001$, $\eta_p^2 = .66$.[2]

---

[2] The shapes of the distributions of Level 3, Level 5, and Level 6 affixes recognised by the learners were not symmetric. Thus, I supplemented the repeated measures ANOVA analysis with a Friedman's test, which does not assume normality. The results of the Friedman's test corroborated the results of the ANOVA confirming that it was robust to the deviations from normality present in the variables, $X^2(3, n = 62) = 114.08$, $p < .001$. The pairwise comparisons also confirmed the results, demonstrating that there were significant differences between all the affix levels except for no difference between Levels 5 and 6.

The effect size value indicated that the affix levels accounted for 66% of all the variance in the learners' performance, which is a very strong effect. The learners' performance is graphically presented in Figure 1. I then compared the means at the affix levels pairwise, in essence conducting a series of posthoc tests using the Bonferroni correction to account for the family-wise error, the results of which are presented in Table 5.
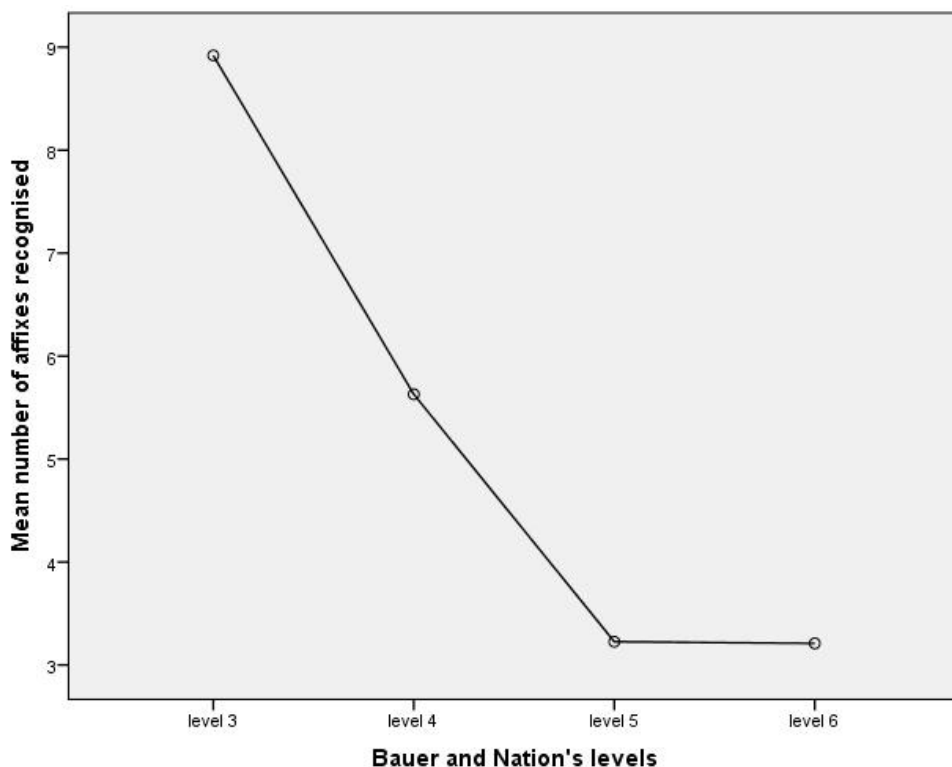

Figure 1 Mean number of affixes recognised at Bauer and Nation's (1993) different levels

Table 5 Pairwise comparisons

Levels	Mean difference	Significance*
Level 3 and 4	3.29	< .001
Level 3 and 5	5.69	< .001
Level 3 and 6	5.71	< .001
Level 4 and 5	2.4	< .001
Level 4 and 6	2.42	< .001
Level 5 and 6	0.02	ns

Notes. *The *p* values were adjusted using the Bonferroni correction; ns = nonsignificant.

The pairwise comparisons, thus, demonstrated that with the exceptions of no significant difference between Levels 5 and 6, all the differences were significant. In fact, the trend was the same if the groups were compared separately, as can be seen in Figure 2.
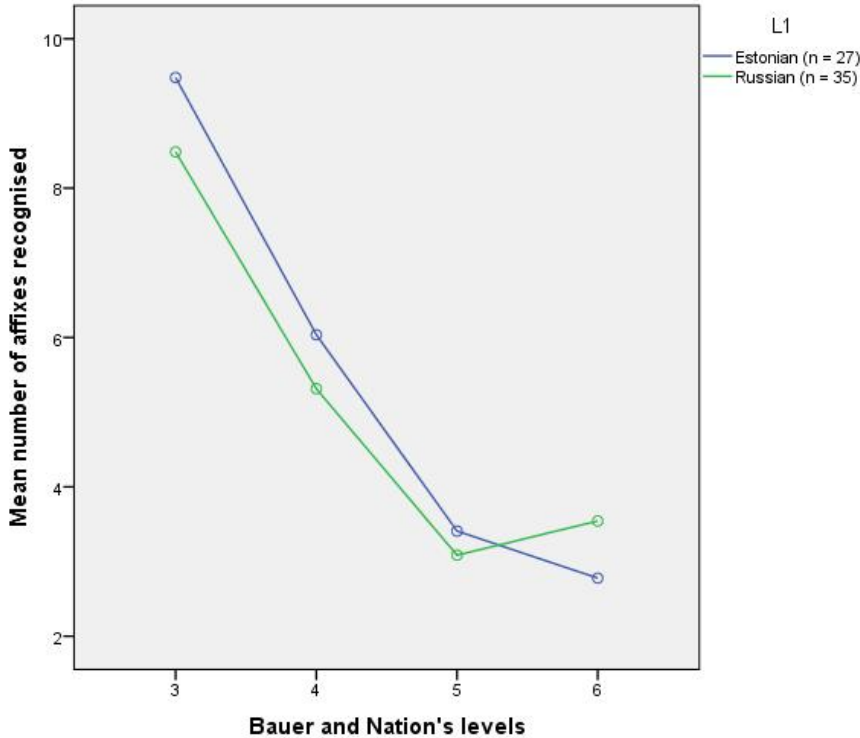

Figure 2 Mean number of affixes recognised at Bauer and Nation's (1993) different levels by L1 Estonian and L1 Russian learners

Moreover, the repeated measures ANOVAs and the pairwise comparisons conducted separately for each of the L1 groups yielded the same results as the analysis conducted for the whole sample. What is more, while, with the exception of the performance on Level 6 affixes, the L1 Estonian group slightly outperformed the L1 Russian group (see Table 4 and Figure 2), there was no significant difference in the ability of either of the two groups to recognise affixes at any of Bauer and Nation's levels, as demonstrated by a series of the independent-samples *t* tests. For example, the greatest mean difference of 0.99 between the performance of the two groups was in the ability to recognise Level 3 affixes, and the *t* test demonstrated that this difference was not statistically significant, $t(60) = 1.18$, $p = .241$.

This being said, the results do not imply that the learners found all the affixes at Level 3 easier to recognise than affixes at Level 4, or all the affixes at Level

6 harder to recognise than Level 4 affixes. For example, many learners recognised the prefix *re-* in *recoup* (45 learners, i.e., 73%) and *reaffirmation* (30 learners, i.e., 48%), which were much higher numbers than those who recognised the rest of the affixes at Level 6 (ranging from 3 to 17 learners). This suggests that the prefix was rather easy to recognise. In fact, these numbers are comparable to those who recognised many of the Level 4 affixes in the task (e.g., the prefix *in-* in *inapt* recognised by 30 learners). Similarly, the numbers of learners who recognised the suffix *-ful* in *pailful* (55 learners, i.e., 89%) and in *boastful* (54 learners, i.e., 87%) were higher than the number of learners who recognised, for example, *-ly* in *stoutly* (51 learners, i.e., 82%), the latter being a Level 3 affix.

What is more, the results do not reveal why substantially different numbers of learners recognised the same affix in different items, such as the prefix *re-*, as illustrated in the previous paragraph. Other examples include the suffix *-ly*, which was recognised in *stoutly by* 51 learners but only by 42 (68%) in *indiscreetly*, or the suffix *-ary* (Level 5), which was recognised by 15 learners (24%) in *deflationary* but only by 2 (3%) in *ternary*.

5. Discussion

The present study aimed at finding empirical evidence for (or against) the order of L2 English affixes proposed by Bauer and Nation (1993). Differently from the previous research, I did not challenge Bauer and Nation's levels by looking at separate affixes but instead considered affixes at different levels as groups. The potential influence of the frequency effect (e.g., Clahsen & Neubauer, 2010) was countered by making sure that the learners did not know the words in which they were asked to find the affixes. Arguably, this allowed for the control of the influence of semantic transparency as well.

The results demonstrated that with the exception of no difference between the learners' ability to recognise Level 5 and Level 6 affixes, the higher Bauer and Nation's level was, the fewer affixes on average the learners recognised at this level. That is to say, their ability to recognise the affixes, for the most part, followed the affix order proposed by Bauer and Nation (1993). What is more, the difficulty order of the affixes accounted for 66% of all the variance in the learners' performance, which is a large effect and should thus be considered a rather strong evidence for the order proposed by Bauer and Nation.

The difference between Level 5 and Level 6 affixes was not statistically significant. The reason for that can, in part, be attributed to the fact that many learners recognised the prefix *re-*. As a matter of fact, Mochizuki and Aizawa (2000) found that the meaning of this prefix (i.e., its most common meaning of *again*) was recognised by the largest number of learners as compared to other

prefixes, which can explain the ease of recognition of the prefix in the present study. What is more, Bauer and Nation (1993) accounted for the undoubted productivity of the prefix, which means that there is a possibility that the learners in the present study met this prefix quite often.

On the other hand, Bauer and Nation (1993) rightfully noted that because of the number of meanings *re-* has in addition to *again* and *anew* and a number of tokens with *re-* that have become lexicalised, learners can end up misanalysing words containing *re-* if they learn the semantics of this prefix. I would, however, suggest that this assumption might be reconsidered in future. *Re-* is present in many languages and its meanings, at least in some of them, are similar to those it has in English. Specifically, it is the case with Estonian and Russian. In Estonian, one can find it, for example, in *representatiivne* ('representative'), where it has an intensifying meaning, and in Russian in *репродукция* [*reproduktsiya*] ('reproduction'), where it has the meaning of *again*. Thus, there could have been the influence of the mother tongue that influenced the learners' performance on the prefix *re-* as well. Further studies, for example, teaching experiments, can shed more light on whether it indeed makes sense to classify the prefix at an earlier stage.

Connected to the previous discussion, it should not be assumed that the learners found all the affixes at Level 3 easier to recognise than Level 4 affixes and all the affixes at Level 4 easier to recognise than those at Levels 5 and 6. One illustration of the opposite is the learners' recognition of *re-* in the task. Another example could be the learners' performance on the items with *-ful*. As regards the latter suffix, it has been found by Mochizuki and Aizawa (2000) that more learners were able to indicate the syntactic role of *-ful* than of suffixes at Level 3, such as *-ly* and *-er*, which can serve an explanation for the finding of the present study. What is more, the meaning of *-ful* also seems easy to remember as it is the same as that of the word *full*, a very frequent word. Presumably, these could be the reasons for the number of learners recognising the suffix in the study.

However, I would refrain from making any claims regarding the potential reclassification of these affixes. The results of the present study do not allow for establishing reasons for the learners' performance on affixes such as *re-* or *-ful*. Moreover, it is hard to say whether comparable numbers of learners recognising the same affixes will be found in future studies. Judging by the previous studies that examined separate affixes and produced dissimilar orders of difficulty, this might not be the case. What is more, while the participants in the present study did not know the meanings of the words in the task, which reduced, if not excluded, the possibility of the effects of frequency and semantic transparency, it is not clear why different numbers of learners recognised the same affixes in different items. I would suggest that before trying to find answers to these questions more research into L2 English word derivation is required, which should

increase our understanding of what is included in L2 English word derivation knowledge and how it develops.

Thus, it is best to interpret the results of the present study in the most straightforward way. That is to say, the results indicate that learners are able to recognise significantly fewer affixes at Bauer and Nation's higher levels than at lower levels; or, perhaps, that learners are more likely to recognise affixes at Bauer and Nation's lower levels than at higher levels.

It should also not be forgotten that I only measured the learners' ability to recognise the affixes. Thus, the results could have been different should I have studied their ability to recognise/recall the meanings of the same affixes, for example.

6. Conclusion

The present study aimed at establishing whether the classification of the English affixes proposed by Bauer and Nation (1993) can indeed account for the difficulty L2 English learners have with recognising derivational affixes. The results demonstrate that with the exception of the lack of the difference between Level 5 and Level 6 affixes, the learners were more likely to recognise affixes at lower levels than at higher levels of difficulty as defined by Bauer and Nation, thus providing evidence for the validity of the levels.

The findings have several implications, both theoretical and practical. The empirical confirmation of the difficulty order reinforces Bauer and Nation's (1993) proposal to use the levels as a reference for affix difficulty in morphological research. Moreover, Bauer and Nation's levels could be used as a starting point for establishing an/the order of acquisition of L2 English derivational affixes, if any. As far as pedagogical implications are concerned, teachers of English could take the levels into consideration when instructing their learners to refer to morphological knowledge when inferring the meanings of unknown words in texts. That is to say, the levels should help L2 English teachers to find which affix properties as defined by Bauer and Nation (1993) make it more likely that their learners will recognise the affixes.

These implications are not new. In fact, Bauer and Nation (1993) discussed these as possible applications of their classification. I, however, argue that the findings reported in the paper present a stronger case for doing so. Having said that, I feel that several limitations of the study should be listed, so that further studies could account for them.

One of the limitations of the study has already been mentioned in Section 3.1. To complete the task, the learners were also required to demonstrate their metalinguistic knowledge, which learners often have problems with (Alderson, Clapham, & Steel, 1997). On the other hand, the word segmentation task is, arguably, the best for determining how well the learners recognise derivational

affixes. I could have rephrased the instructions, and, similarly to Hayashi and Murphy (2010), instead of mentioning prefixes and suffixes in the instructions, asked the learners to break the words into meaningful units. This, however, would have meant that the learners would also have to identify the bases, and I wanted to minimise this possibility. What is more, if the instructions had been phrased without mentioning prefixes and suffixes, the learners might have misinterpreted what they were asked to do (cf. Hayashi & Murphy, 2010).

Another limitation concerns the inability to say whether the results would be exactly the same if other affixes had been used. Using other, perhaps more easily recognisable affixes at Level 5, such as *-hood*, *post-*, or *neo-*, could result in learners recognising more affixes at Level 5. Finally, the way I controlled for semantic transparency might have been insufficient. Further studies, using other affixes, and, perhaps, a larger number of affixes as well as a better control for semantic transparency could confirm or disprove the findings of the present study. Furthermore, future studies could also determine whether learners find it easier to recognise or recall meanings and/or syntactic roles of affixes at Bauer and Nation's different levels taken as a group. This would strengthen the case for Bauer and Nation's levels or present evidence against them. In any case, I hope that the present study stimulates the research on L2 English word derivation knowledge.

References

Alderson, J. C., Clapham, C., & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research, 1*(2), 93-121.

Alegre, M., & Gordon, P. (1999). Rule-based versus associative processes in derivational morphology. *Brain and Language, 68*(1-2), 347-354.

Bailey, N., Madden, C., & Krashen, S. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning, 24*(2), 234-243. doi: 10.1111/j.1467-1770.1974.tb00505.x

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing, 12*(3), 169-190. doi: 10.1023/A:1008131926604

Carlisle, J. F., & Fleming, J. (2003). Lexical processing of morphologically complex words in the elementary years. *Scientific Studies of Reading, 7*(3), 239-253. doi: 10.1207/S1532799XSSR0703_3

Chuenjundaeng, J. (2006). *An investigation of SUT students' receptive knowledge of English noun suffixes* (Unpublished MA thesis). Suranaree University of Technology, Thailand.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning, 60*(1), 21-43. doi: 10.1111/j.1467-9922.2009.00550.x

Clahsen, H., & Neubauer, K. (2010). Morphology, frequency, and the processing of derived words in native and non-native speakers. *Lingua, 120*(11), 2627-2637.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment* [electronic version]. Retrieved from http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf

Felser, C., & Clahsen, H. (2009). Grammatical processing of spoken language in child and adult language learners. *Journal of Psycholinguistic Research, 38*(3), 305-319. doi: 10.1007/s10936-009-9104-8

Friedline, B. E. (2011). *Challenges in the second language acquisition of derivational morphology: From theory to practice* (Unpublished doctoral dissertation). University of Pittsburgh, USA.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 51*(1), 1-50. doi: 10.1111/1467-9922.00147.

Gollan, T. H., Foster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English

bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 112-239.

Hayashi, Y., & Murphy, V. (2010). An investigation of morphological awareness in Japanese learners of English. *The Language Learning Journal, 39*(1), 105-120. doi: 10.1080/09571731003663614

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21*(1), 47-77. doi: 10.1093/applin/21.1.47

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*(4), 603-634. doi: 10.1017/S0142716404001298

Lardiere, D. (1998). Case and tense in the 'fossilized' steady state. *Second Language Research, 14*(1), 1-26. doi: 10.1191/026765898674105303

Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning, 26*(1), 125-134.

Leontjev, D. (2016). *ICAnDoiT: The impact of computerised adaptive corrective feedback on L2 English learners* (Unpublished doctoral dissertation). University of Jyväskylä, Finland.

Marslen-Wilson, W. (2007). Morphological processes in language comprehension. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 175-193). Oxford: Oxford University Press.

Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System, 28*(2), 291-304.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Perlmutter, D. (1988). The split morphology hypothesis: Evidence from Yiddish. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology* (pp. 79-100). San Diego: Academic Press.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.

Põhikooli riiklik õppekava õigusakt; Lisa 2 [Basic School National Curriculum Act: Annex 2.] (2010). Retrieved from https://www.riigiteataja.ee/aktilisa/1281/2201/0017/13275423.pdf

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition, 19*(1), 17-36.

Schmitt, N., & Zimmermann, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145-171. doi: 10.2307/3588328

Schreuder, R. & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131-156). Hillsdale, NJ: Lawrence Erlbaum.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition, 11*(2), 245-260. doi: 10.1017/S1366728908003404

Thorndike, E. L. (1942). The teaching of English suffixes. *Teachers College, 43*(8), 657-658.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1-2), 231-270.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition.* Norwood, NJ: Ablex.

APPENDIX

The word segmentation task (originally, the instructions were in the participants' mother tongues; the correct responses are highlighted)

Some of the words you see below are formed with help of prefixes OR suffixes OR both prefixes and suffixes. Circle the prefixes and suffixes in these words.
In some words there are both a prefix and a suffix. In some words, there are only prefixes. In some, there are only suffixes. In some of the words, there is neither a prefix nor a suffix.
If you know the meanings the words, write what they mean (a translation or a definition). But even if you don't know the meaning of any of the words, don't worry. These are very difficult words. Even if you know the meanings of one or two, you have got a rather good vocabulary.

The first two words are examples.

FARM(ER)	_____talunik_____
(DE)ASH(ED)	_____
unshackle	_____
mediocre	_____
pailful	_____
slanderous	_____
recoup	_____
briskness	_____
unambiguous	_____
indiscreetly	_____
voidable	_____
comprise	_____
heraldic	_____
lushness	_____
regressive	_____
repentant	_____
decipherable	_____
eviction	_____
blandness	_____
stoutly	_____
frugality	_____
crofter	_____
prohibitive	_____
deflationary	_____
inapt	_____
brimless	_____
enshrinement	_____
abolish	_____
boastful	_____
discernment	_____
obstructive	_____
bulletin	_____

interlace	_____
solemnise	_____
errant	_____
mortify	_____
cherubic	_____
moisten	_____
obscenely	_____
digression	_____
magnitude	_____
scrutiny	_____
discipleship	_____
deference	_____
detainee	_____
ternary	_____
reaffirmation	_____
enmesh	_____
arsonist	_____
bestowal	_____
exemption	_____
misapprehend	_____