Author(s): Olkkonen, Sanna

Title: Processing limitations in L2 fluency : Analysis of inaccuracies in lexical access

Year: 2017

Version:

Please cite the original version:

Olkkonen, S. (2017). Processing limitations in L2 fluency : Analysis of inaccuracies in lexical access. Apples : Journal of Applied Language Studies, 11(1), 19-41. https://doi.org/10.17011/apples/urn.201704252081

Apples – Journal of Applied Language Studies
Vol. 11, 1, 2017, 19–41

Processing limitations in L2 fluency: Analysis of inaccuracies in lexical access

Sanna Olkkonen, University of Jyväskylä

Oral fluency is widely included in second language assessments, but its relationship to language proficiency is not straightforward. In the current study, data gathered in an experimental setting were examined with an exploratory fluency analysis. The aim was to examine the relationship between fluency of lexical access and proficiency in foreign language (L2). Fluency of the lexical access was studied by analysing inaccuracies in one word recognition and one word retrieval task. To see if proficiency had an effect on the number and the type of inaccuracies, lexical access tasks were carried out for 563 Finnish school children from grades 4, 8, and 11 in their L2 (English). Proficiency in L2 was expected to develop during school education. The inaccuracies were proposed to stem from processing limitations in language use, i.e., inefficiency of lexical access, or from control of attention. The hypothesis was that if lexical access is not automatized, there are less resources for attention-control in recognising and retrieving words. Therefore, the inaccuracies in L2 relating to inefficiency were hypothesised to decrease with proficiency, whereas the ones relating to control of attention were proposed to be more stable or to increase. Furthermore, the fluency of L1 lexical access was used as a control measure. The results offered some confirmation to these hypotheses. For example, some evidence for more available resources in correcting and monitoring speech was found for the older students. The overall results highlight caution in assessing L2 fluency, as not all types of inaccuracies were connected with lower proficiency.

Keywords: fluency, oral proficiency, information-processing, lexical access

1 Introduction

Second language (L2) fluency has received growing interest in recent years (e.g., Akamatsu, 2008; Bosker, Pinget, Quené, Sanders & De Jong, 2013; De Jong, Groenhout, Schoonen & Hulstijn, 2015; Engelhardt, Corley, Nigg & Ferreira, 2010; Housen & Kuiken, 2009; Kahng, 2014; Kame'enui & Simmons, 2001; Riggenbach, 2000; Segalowitz, 2010; Snellings, Van Gelderen & De Glopper, 2002). No clear consensus has been reached, though, on what is meant by fluency of language use. Despite this, oral fluency is widely used in assessing L2 proficiency; it is, for example, one of the five assessment criteria in the Common European Framework of Reference for Languages (CEFR, Council of Europe, 2007; see also Koponen & Riggenbach, 2000). The criteria given for assessing fluency are very wide

Corresponding author's email: sanna.olkkonen@jyu.fi
ISSN: 1457-9863
Publisher: Centre for Applied Language Studies
University of Jyväskylä

http://apples.jyu.fi
http://dx.doi.org/10.17011/apples/urn.201704252081

generalisations, and CEFR defines spoken fluency in terms of pausing and 'natural, smooth flow of language'. This is problematic, as it is questionable whether even L1 speech can be considered as 'naturally smooth' when it includes pausing, self-corrections, and repetitions as well (Lennon, 1990). Furthermore, not all of the aspects of oral fluency seem to be as indicative of proficiency level as others. As De Jong and colleagues (2015) showed, the duration of pauses is more closely connected to L1 speaking style than to L2 proficiency. Therefore, duration of pauses should be considered only to a modest extent when assessing L2 fluency. This highlights the importance of understanding, which are the features of oral fluency that are most reliably indicative of L2 proficiency.

In the current study, it is proposed that oral fluency and its relationship to L2 proficiency might be studied from the viewpoint of processing limitations. For the beginning language learners, even the most basic aspects of language, such as phonological coding and lexical search require much of the resources. When these processes start to automatize with proficiency, the resources can be directed to higher aspects of language, such as more global planning and comprehension monitoring. Fluent language use thus means a balance between the *automatic* and *controlled* processes (Schmidt, 1992; Segalowitz, 2000, 2010, p. 91). Examining these two levels of processing in L2 use may reveal us important information on how fluency and proficiency are connected. Therefore, it is proposed that the *inaccuracies* in language production may offer a fruitful way of tapping into these processes. Fluency was measured with two standardised lexical access tasks (measuring word recognition and word retrieval), and in addition to L2, these tasks were conducted also in L1 to use its fluency as a control variable. To be able to examine the differences between proficiency levels, the tasks were conducted for 580 Finnish schoolchildren from three different L2 proficiency levels: Grade 4 (in primary school, age 10), Grade 8 (in lower secondary school, age 14), and Grade 11 (in upper secondary school, age 17). The use of grade levels as a proxy for the developing proficiency has been previously confirmed with L2 writing tasks, assessed along the CEFR 6-point scale (see Olkkonen, Eklund & Leppänen, submitted).

1.1 Defining and measuring fluency

The range of definitions of L2 fluency is wide. It can mean the overall proficiency of a language; only the speed of language production; the ease or efficiency of production; or including prosodically appropriate expression and phrasing (for overviews: Housen & Kuiken, 2009; Koponen & Riggenbach, 2000). The most comprehensive model to date, however, is Segalowitz's (2010) distinction between cognitive, utterance, and perceived fluency. The model builds on Lennon's division (1990) of speech production and underlying psycholinguistic processes, where the surface fluency is a reflection of the underlying cognitive fluency. *Perceived fluency* refers to the interpretations the listeners make on a speaker's proficiency based on their surface fluency (cf. Bosker et al., 2013). *Utterance fluency* refers to the surface structure that reflects the cognitive level and is directly measurable. Utterance fluency is often operationalised as the rate of repairs, pauses, and speed (Skehan, 2003). Pausing and speed have been shown to be relatively good markers of (dis)fluency (Bosker et al., 2013; De Jong et al., 2015; Freed, 2000; Kahng, 2014). On the other hand, the repairs (false starts, repetitions, replacements, and reformulations) have been studied to a much lesser extent, and their relationship to L2 proficiency has been questioned (Engelhardt et al., 2010;

Gilabert, 2007; Kormos, 1999; Lennon, 1990). These results have important implications for the current study as well, when considering the different types of disfluencies (inaccuracies) and proficiency.

Cognitive fluency refers to how efficiently (i.e., quickly and accurately) the processes underlying language production operate (Segalowitz, 2010, p. 48). These underlying processes include decoding words and lexical access (Grabe, 2009). Our cognitive resources are limited, and thus, the more fluently these underlying processes operate, the more smoothly the higher-level processes can be executed (Kirby, Parrila & Pfeiffer, 2003; LaBerge & Samuels, 1974). These higher-level processes include, e.g., monitoring and comprehension. In Segalowitz's formulation, cognitive fluency consists of two complementary processes: efficiency and attention-control (2000, 2010, p. 91). Efficient language use (whether reading, speaking, writing or listening) entails fast and accurate processing. When this processing is inefficient, there are fewer resources available for the attention-related processes (Kahng, 2014; Kormos, 2000; Walczyk, Marsiglia, Johns & Bryan, 2004). This means, for example, that a speaker is not able to efficiently suppress irrelevant information, shift between different requirements of the situation, or monitor one's performance (see Segalowitz, 2010, p. 93). These skills are very important for language use, as Segalowitz and Frenkiel-Fishman (2005) showed that the efficiency of L1 and L2 task-switching accounted for 59% of the variance in L2 proficiency.

Studies aimed at cognitive fluency measurements are rare, and often highly indirect measures are used, such as stimulated recall settings (Kahng, 2014). One possibility for a more direct measurement has been proposed to be the efficiency of access to lexicon (Segalowitz, 2010, p. 75; Segalowitz & Frenkiel-Fishman, 2005). Even though often measured orally, the efficiency of lexical access has been shown to be a good reflection of fluency of mental processing (Denckla & Cutting, 1999; Di Filippo et al., 2005). Compared to longer monologue or dialogue settings, the reduced contextual and communicative aspects of the lexical access measurements may help to tap into the cognitive processes in a more ecologically valid way. The term *lexical access* has been used as both referring to word retrieval (e.g., Gholamain & Geva, 1999; Kirby et al., 2003), and to word recognition (e.g., Akamatsu, 2008; Grabe, 2009). On the other hand, they might be both interpreted as related processes, both tapping into lexical access but from different directions (Snellings et al., 2002). Both word retrieval and recognition rely on retrieving a phonological form of words, while utilising different subcomponents of lexical access (see also Gollan et al., 2011). In the current study, this approach was applied by measuring fluency in both word recognition and retrieval.

1.2 Inaccuracies and processing limitations in lexical access

Lennon, in his account on L2 fluency, stated that "for most speakers in most situations, processing demands, rather than deficient knowledge, will limit fluency" (2000, p. 27). Therefore, studying the connections between L2 fluency and proficiency might be performed by tapping the processing demands. Especially, the processing limitations in language use may prove a fruitful source of information. As Just and Carpenter's (1992) influential theory of information processing defines, human cognition is limited and these limits are set by the working memory capacity. When our language skills are in development, the cognitive resources are directed by necessity to low-level processes, such as lexical access. There is not enough capacity for the higher processes, like comprehension or correcting your speech, which may result as errors in language production. These errors, therefore, serve as reflections of the proficiency level. On the other hand, Ehri (1991) has hypothesised that certain types of errors could be connected with the development of language skills, so that they appear only after certain proficiency threshold. Therefore, the breakdowns of fluency may also be indications of developing proficiency, and, as such, they can offer important insights into the influence of processing capacity in the language use (see e.g., Engelhardt et al., 2010; Protopapas, Fakou, Drakopoulou, Skaloumbakas & Mouzaki, 2013).

Considering the terminology, the term *error* has been used in accuracy-oriented literature (Di Filippo et al., 2005; Ehri, 1991; Geva & Siegel, 2000; Gilabert, 2007; Kormos, 2000) and it implies a dichotomical distinction between correct and incorrect answer. This does not seem to cover sufficiently the model applied in the current study in the case of, e.g., self-corrections where speakers consciously and actively modify their product. The term *disfluency*, on the other hand, is used in the utterance fluency research to refer to surface properties of fluency and it is considered separate from accuracy (e.g., Bosker et al., 2013; Kahng, 2014). In the current paper, the term *inaccuracy* is therefore used, both to avoid confusion with the other theories' operationalisations and, further, to imply that the categorisation used in the current paper includes both the breakdowns relating to inefficiency and to control of attention.

Very few studies have measured inaccuracies of lexical access, and mostly only the percentage of correct answers has been of interest (e.g., Akamatsu, 2008; Gholamain & Geva, 1999; Salmi, 2008). Operationalisation or theorisation of inaccuracies in lexical access is nearly non-existent, and detailed analyses are rarely considered. One study attempting to categorise the inaccuracies in lexical access is Kaukonen and Lanu (2005), in which the authors analysed both weak and normal readers' word reading errors in L1 Finnish. They found the most frequent error types to be guessing, difficulties in pronunciation, word recognition problems, and speed–accuracy trade-off. These error types were found to differentiate the weak and normal readers significantly. Based on these results, the authors hypothesised a model for inaccuracies stemming from the undeveloped automaticity, working memory restrictions, lexical restrictions, motivation problems, and problems in maintaining alertness. Danielsson (2003) conducted a similar study in L1 Swedish, explaining the reasons for errors in word reading mainly on orthographic or linguistic principles (see also Geva & Siegel, 2000). These studies have not, however, considered the categorisations in relation to processing limitations. According to Segalowitz's cognitive fluency model (2010), the errors might be interpreted as stemming from either the inefficient language skills or from the attention-related processes. Next, I will discuss how these might be realised in lexical access.

1.2.1 Inefficient processing

First, cognitive fluency entails efficient, i.e., both fast and accurate processing. For example, to be a fluent reader, most words in the text have to be recognised as wholes (Ehri, 1991). If the language skills are limited, the speaker may have to choose between speed and accuracy, which leads to trade-off (Grabe, 2009, p. 292; Kame'enui & Simmons, 2001; Walczyk et al., 2004). Favouring speed over accuracy may lead to, for example, *skipping* items, whether by accident or for buying time (cf., Marian, Blumenfeld, Mizrahi, Kania & Cordes, 2013). When aspiring to read quickly, the readers may also revert to *guessing* words, especially when encountering low-frequency words (see Balota et al., 2006, for an overview). The low-frequency words are often read as more frequent items of similar length and overall graphemic structure, i.e., visual form (Balota, Yap & Cortese, 2006; Broadbent, 1967; Danielsson, 2003). Similarly, Ellis (2002) notes that in listening, more common words are perceived correctly more quickly, and incorrect responses usually stem from a small set of relatively common words. Cossu, Shankweiler, Liberman, and Gugliotta (1995) did not find visual form affecting misreadings of individual letters as much as phonological interference (b–d and d–t but not u–n), but the case might be different for whole words (e.g., word superiority effect, see Balota et al., 2006). Gollan and her colleagues (2011) found that word frequency was more closely related to word recognition than to naming speed (example of word retrieval), which involves only highly familiar words. Inefficiency thus manifests itself most likely in different ways in word recognition and retrieval tasks, and in the current setting, it was possible to compare if this was the case.

1.2.2 Attentional processes

Attention-control may also induce breakdowns of fluency, e.g., in the form of repairs (Skehan, 2003). *Correcting and monitoring* one's own production disrupts the flow of speech, but also requires cognitive resources. For the beginning language learners, there may not be enough resources to allocate to these higher processes (cf., Kahng, 2014; Kormos, 2000). In the previous literature, self-corrections have been hypothesised to be more prominent amongst the more proficient language learners but empirical findings are scarce (Lennon, 1990; Segalowitz, 2010; cf., Danielsson, 2003). In Kaukonen and Lanu's (2005) data, the number of self-corrections did not decrease from second to third grade for either the normal or weak readers. Freed (2000) found that students with more advanced speaking skills (stay-abroad experience) attempted to correct their expressions and to reformulate their speech to a greater extent than their less advanced peers, thus producing more false starts. Repetitions and self-corrections, further, did not differentiate between the at-home and abroad groups. For *repetitions*, Peltonen and Lintunen (2016) found that these were more connected with personal speaking styles and strategies than inadequate language skills in their L2 speech data. Repetitions showed very much within-group variation and their use did not mostly differentiate between proficiency groups (see also Bosker et al., 2013). This implies that speakers may use repetitions in a functional manner, to avoid excessively long silences, to keep the speech-turn, and to minimise disruptions in the speech (Peltonen & Lintunen, 2016). The use of repetitions, furthermore, did not differentiate the ADHD and normal control groups in a sentence production task (Engelhardt et al., 2010). In the study by Kahng (2014), the repetitions correlated only weakly with speaking scores and corrections did not correlate at all, although the L2 speakers used these repairs more than the L1 speakers did. The role of the repairs in language proficiency remains thus quite unclear. In the current study, it is hypothesised that these might be related to more advanced language skills, when there are more resources to apply them.

Attention-control entails, furthermore, *inhibition or suppression of irrelevant information*. The difficulty in inhibiting irrelevant information is very often illustrated by the Stroop task (Stroop, 1935), where participants are instructed to name the colour of the print instead of reading a colour word. Naming times are longer when the colour and word are incongruent (e.g., word "RED" written in blue ink), and this indicates that reading of the colour name has to be suppressed in order to be able to name the ink (for a review, see MacLeod, 1991). It has been shown that, with better language skills, it is in fact more difficult to inhibit irrelevant material. Automatic activation of the word meaning is stronger in L1 than in L2, for example, and suppressing this activation requires cognitive resources (Marian et al., 2013; see also Gernsbacher, 1993). Favreau and Segalowitz (1983) showed that less-skilled L2 users benefited from a priming effect (e.g., were quicker to recognise the word "apple" when preceded by "fruit" rather than by "furniture") only when there was enough time to consciously process the material (1150ms). More skilled L2 users showed a priming effect even with shorter exposure times (200ms). This illustrates how automatic processing, and the lack thereof, depends on the language proficiency, and this affects also the inaccuracies that may be produced.

The lexical access tasks used in the current study (single word recognition and word retrieval) lack the contextual clues for priming to happen, but difficulties in suppressing irrelevant material may result in inaccuracies stemming from successive words in the list format. Eye-tracking studies offer support for the view that automatization makes inhibiting the irrelevant material more difficult. Fast readers are found to pre-process words just outside of the fixation point, and this has been shown to influence processing of the fixated word (Pollatsek, Rayner & Balota, 1986; Simola, Holmqvist & Lindgren, 2009; shown also in listening tasks: Ellis, 2002). Jones, Obregón, Kelly, and Branigan (2008) found that even in a naming speed task (RAN), there was sensitivity to some visual and phonological information from material in the parafoveal preview (i.e., material not yet focused on). This is an indication of the automatic connections to lexical storage, which are activated without conscious effort, but require higher levels of language proficiency. Therefore, it is hypothesised here that the beginning language learners do not encounter inhibition difficulties. In reading a word list and naming items, some pre-processed information from the visual field may influence the currently produced items of the more advanced L2 learners; however, they might follow a u-shape thus that the most advanced students have enough resources for suppressing to happen appropriately.

Switching between different tasks or mental sets also requires attentional capacity. This has been studied, e.g., with number-letter task where participants have to decide whether, in a number-letter combination (e.g., 7G), the given number is even or odd, or the given letter is a consonant or vowel, depending on the combination's location on a screen (Miyake et al., 2000). Shifting, when including two different languages, has been proposed to be easier from L1 to L2 than vice versa (cf., Meuter & Allport, 1999). This is because L1 needs to be heavily suppressed to be able to perform in the less automatized L2, and Segalowitz and Frenkiel-Fishman (2005) report that there is more shift-cost when switching from L2 to L1 than vice versa. As was proposed with inhibition, it may be that it is more difficult to suppress a previously activated category with more automatized language skills than in a language that is still less developed and with lower activation level. The current study explored the effects of attention-control in lexical access tasks not only between different-level learners, but also in both L1 and L2 to find possible differences between more and less automatized languages.

1.3 Aims of the present study

The current study concentrated on a categorisation of inaccuracies in lexical access. The categories were based on processing limitations in language production. The sources of inaccuracies were considered to be *inefficient processing* and *control of attention*. The hypothesis was that if lexical access is not automatized, there are no resources for attention-control. Only when proficiency develops, can cognitive resources be allocated to monitoring own production. The role of proficiency was examined, first, by the types of inaccuracies between the L2 learners of three grade levels. Second, the inaccuracies between the L1 and L2 results were compared within the grades to see if the number and the type of the inaccuracies differed according to more or less proficient languages.

2 Data and method

The data were gathered in an experimental, cross-sectional setting, with participants from three different grades as proxies for L2 proficiency levels. To measure fluency of lexical access, two standardised psycholinguistic tasks were used, originally designed for L1 diagnosis. Here, they were conducted in L2 to study their interaction with L2 proficiency in an experimental manner. In contrast to the usual measurement of speed and accuracy, here a more exploratory categorisation of inaccuracies was used. Furthermore, fluency of L1 lexical access was measured as well for comparison data. The data were analysed quantitatively by counting the number and proportions of different inaccuracies and comparing them between the grades. The aim was to see if there were differences in the occurrences of the types of inaccuracies between the proficiency levels. In addition, χ^2 tests were used to examine if any of the possible differences between the proficiency groups were statistically significant. The inaccuracies were further explored comparing the performance in L2 lexical access tasks to the similar ones in L1, to see if there were differences in the less and more proficient languages within the grade levels.

2.1 Participants

The participants (N=563, 53% females) were Finnish schoolchildren who studied English as a foreign language. There were three age groups: *Grade 4* (age = 10, n = 192), *Grade 8* (age = 14, n = 186), and *Grade 11* (school year 11, age = 17, n = 185). All the participants had learned English since Grade 3, and the grade levels were used as an approximation of the L2 level (as explained further in Olkkonen et al., submitted). As Finnish orthography is very transparent, children are mostly very fluent readers of L1 by Grade 4 (see e.g., Eklund, Torppa, Aro, Leppänen & Lyytinen, 2015). The data were gathered in DIALUKI project in Finland from 2010 to 2011. The students came from 37 different schools around Finland, in both rural and urban areas. The testing was conducted during the school hours and consents to participate in the study were obtained from the counties, the schools, the parents, and the children themselves.

2.2 Procedure and materials

The lexical access tasks were a part of a cognitive test battery (appr. 45 minutes), that was conducted individually in a quiet room by trained assistants with a *Cognitive Workshop* software (the Finnish version developed by the University of Dundee and the Jyväskylä Longitudinal Study of Dyslexia, see Lyytinen et al., 2004; the English versions by DIALUKI project). Two lexical access tasks were used: one word retrieval (RAS) and one word recognition (Word List), with similar tasks in both L1 and L2. The order of administering the L1 and L2 versions was counterbalanced to minimise the language order effect.

Rapid Alternating Stimulus (RAS) task was used for measuring naming speed and retrieval of familiar vocabulary (Wolf, 1986). In L1 for all groups and in L2 for Grade 8 and 11, the matrix consisted of 50 units, including numbers (2, 6, 7, and 9), colours (black, blue, green, red, and yellow), and letters (A, E, S, and T), arranged in five rows in a semi-random order (from Ahonen, Tuovinen & Leppäsaari, 2003). For Grade 4, the task was slightly easier and shorter in L2 and it consisted of 30 units, with numbers (2, 6, 7, and 9), colours (black, blue, green, red, and yellow), and familiar pictures (pencil, car, fish, and house). Picture naming was considered to be more automatized than foreign language letter naming at this level (see Denckla & Cutting, 1999). In comparison to simple RAN, which uses items from only one semantic category, RAS requires shifting between different semantic categories that makes it a more challenging task (see Wolf, 1986). Before the task, the assistant checked that the child knew the names of all the used items. Participants were asked to name the items aloud as fast and as accurately as they could, and the time to conduct the task was measured by a stopwatch. The task was conducted individually for each student and recorded to the computer with a microphone-headset for later analyses.

Word List reading. The children were given a printed word list of 105 words, with the words arranged in three columns, to be read from top to bottom. In the beginning of the list, the words were short and familiar, gradually becoming longer and more complex. Children were instructed to read aloud as many words as they could in 60 seconds' time, as fast and as accurately as possible. The task was conducted individually, in both L1 and L2. The task originated from the standardised Lukilasse reading test battery for L1 Finnish in primary school (Häyrinen, Serenius-Sirve & Korkman, 1999). The English word list was designed similarly in DIALUKI project by sampling words from a frequency list[1]. Even though the number of words was equal, the lists differed in the number of syllables (Finnish 379 syllables; English 183 syllables). This was due to affixation and complex inflectional morphology of the Finnish language: the longest words to be read in L1 were 22 letters long (e.g., *prosessikirjoittaminen* 'process writing' versus L2 English ten letters in *experience*). This made the Finnish word list quite challenging; however, as the phoneme-grapheme-connections are very reliable and the orthography is transparent, the decoding of the words is quite easy even for young readers (Aro & Wimmer, 2003).

2.3 Classification of inaccuracies in lexical access

To assess the question of the types of inaccuracies in relation to cognitive fluency, a classification of the inaccuracies was devised, based on the Segalowitz's (2000, 2010) fluency construct. The inaccuracies were proposed to stem from two sources, inefficiency and attention-control. First, the inaccuracies that relate to inefficiency were hypothesised to be visible with lower-proficiency L2 levels, and to decrease when proficiency increases. Second, the control of attention was proposed to show in more proficient language-learners who have more resources for the controlled processes, but may also encounter more involuntary activation of irrelevant material. All the responses in the RAS and Word List tasks were analysed according to the classifications described below (summarised in Table 1.). The inaccuracies were coded from the sound files that were recorded at the testing situation. The number of participants included in each analysis differed slightly between the tasks due to technical problems in the recordings. One item could include several inaccuracies (e.g., correcting a skipped item). As the classifications were exploratory and without clear correct / incorrect dichotomy, 10% of the data from each grade were double-checked from the sound files by a second rater, informed on the original assessment criteria. The overall agreement rate on accuracy, and place and type of inaccuracies was 95.6%.

a) Inefficient processing
   The inefficiency inaccuracies were proposed to stem from non-automatic lexical access, where processing capacity was required for recognising and retrieving words. This was assumed to relate to lower proficiency of L2 skills: i.e., to be found more for the beginning language learners and to decrease with more proficiency. In addition to the possible categories discussed in Chapter 1.2, errors within the same category and pronunciation difficulties were included.

1. category *Guessing (GUE)*. Fast performance requires sight-word reading (Ehri, 1991), and when word recognition is not automatized this can lead to guessing words. In the RAS task, this category included visually similar items, for example the letter '*A*' mistakenly named as the number '*4*'. In word recognition, guessing often follows a frequency-based bias, which results in inaccuracies that are graphemically similar, but more frequent ones than the target words (Broadbent, 1967; Balota et al., 2006). This category of inaccuracies was hypothesised to decrease as proficiency increased. The category included only real words (Example 1) and possible, readily understandable neologisms (Example 2); nonwords were not included (similarly as Danielsson, 2003).

Example 1

Finnish	heilahdella 'swinging' > heilahdus 'a swing'
	turkikset 'furs' > turkkilaiset 'Turkish'
	Ilmarinen (character in Finnish mythology) > ilmainen 'free', imarrella 'to flatter'
	haluttaisiin 'would be wanted' > hautajaisiin 'to the funeral'
English	probable > probably, course > of course
	place > please, move > movie, find > Finland

Example 2

Finnish	professori 'a professor' > professoida 'to (act like a) professor'
	kierrätyskeskus 'recycling centre' > kierrähdyskeskus 'revolving centre'
English	particular > practicular

2. category *Skipping (SKIP)*. Inefficiency of lexical access was further proposed to be visible in skipping individual words in Word List or items in RAS, whether purposefully or accidentally, when aspiring for speed (cf., Marian et al., 2013).
3. category *Same-category (SCAT)*. Inefficient lexical access in RAS task was interpreted to show in inaccuracies within the same category (e.g., blue > green). The requirement for speed was proposed to cause inaccuracies even with very automatized material. The results were controlled for the possible colour-blindness.
4. category *Pronunciation problems (PRON)*. This category included miscellaneous problems in pronouncing items, e.g., reading English words according to Finnish grapheme-to-phoneme rules, and otherwise unclassifiable instances of pronunciation problems.

b) Attentional processes
Inaccuracies relating to control of attention were hypothesised to increase with proficiency, relating to two trends. First, with higher proficiency, there are more resources available for monitoring own speech as the lower-level language processes are becoming automatized. On the other hand, with higher proficiency, there is also a possibility of automatic activation of lexical material that is involuntary and can disturb the production.

5. category *Self-correction (SCOR)*. Correcting one's own production was considered to require processing capacity and to increase with increasing proficiency of language skills (Walczyk et al., 2004; Segalowitz, 2010; also Schmidt, 1992). No further distinctions were drawn on these, and whether the correction was successful or not was not included in the analyses (cf., Kormos, 2000).
6. category *Repetition (REP)*. Repeating items is included in the repair phenomena of speech (Skehan, 2003), and therefore, it was considered here to be a part of conscious monitoring of speech. Repetitions are proposed, furthermore, to be possible strategic devices and an intentional way of buying time during speech (Peltonen & Lintunen, 2016). In the current study, this category only included repeating whole items, whereas partial repetitions were included in pronunciation problems (cf., De Jong et al., 2015).
7. *Inhibition difficulties (INH)*. As the processing of words begins slightly before the gaze is fixated on them (Jones et al., 2008; Simola et al., 2009), succeeding material can affect performance especially in time-pressure situations. This is shown in the Example 3 where a plural ending *–eet* from word number 43 is also copied to the word number 42. This was hypothesised to be an indication of more fluent reading, and involuntary parafoveal word processing.

Example 3

Finnish 42. kyynel > kyynel/eet 'tear > tear/s'
43. pyyhk/eet 'towels'

8. *Shifting difficulties (SHIF)*. In the RAS task, the difficulty of suppressing an activated category was proposed to cause inaccuracies in category shifting. An example of category shifting difficulty is when the number *six* is followed by the letter '*S*', and is, therefore, named to be the number *seven*. The strength of activation is proposed to depend on the proficiency level, so that

with more automatic language skills the activation is also stronger (Segalowitz & Frenkiel-Fishman, 2005).

Table 1. Summary of the sub-categories of inaccuracies in lexical access and their descriptions.

Category	Description
Inefficiency	
1. Guessing (GUES)	producing wrong responses that are graphemically similar to the target words, frequency-based (Broadbent, 1967; Balota et al., 2006)
2. Skipping (SKIP)	skipping items either accidentally or strategically (Marian et al., 2013)
3. Same-category (SCAT) * only in RAS task	producing wrong responses that are within the same category, such as naming 6 as 7, A as E, or green as blue
4. Pronunciation problems (PRON)	stammering when encountering difficult phonemes or combinations
Attention	
5. Self-correction (SCOR)	correcting one's own speech requires attention and cognitive resources (Segalowitz, 2010)
6. Repetition (REP)	repeating items, subject to individual differences and strategic use (e.g., Kahng, 2014; Peltonen & Lintunen, 2016)
7. Inhibition difficulties (INH)	difficulty with suppressing irrelevant information are proposed to be connected with more advanced skills as the automatic activation is stronger (Marian et al., 2013; Jones et al., 2008)
8. Set-shifting (SHIF) * only in RAS task	set-shifting is more difficult with stronger activation and may be easier with less proficiency; e.g., switching from activated letters to numbers (Segalowitz & Frenkiel-Fishman, 2005)

3 Results

Overall, the accuracy in both of the lexical access tasks was very high. Accuracy percentages for English L2 RAS were 96% (Grade 4), 93% (Grade 8), and 97% (Grade 11). For Finnish L1 RAS, the accuracy percentages were 96% (Grade 4), 98% (Grade 8), and 99% (Grade 11). The accuracy percentages for English L2 Word List were 88% (Grade 4), 95% (Grade 8), and 97% (Grade 11). To compare, the Finnish L1 the accuracy percentages for Word List were 89% (Grade 4), 94% (Grade 8), and 96% (Grade 11; for further discussion on the overall accuracy in the tasks, see Olkkonen et al., submitted). The number of inaccuracies was, therefore, small, as most of the items named or read were correct and fluent. Nonetheless, the distributions of inaccuracies can still offer some insights into differences between the proficiency levels, even though the distributions were mostly too skewed for rigorous statistical analyses.

Thus, first relative proportions of inaccuracies were counted to compare the performance *between* the grades, in both languages. The number of items produced was different in both L2 RAS, where the number of items differed between Grade 4 (30 items) and the higher grades (50 items), as well as in L1 and L2 Word Lists, where the number of words read varied for each individual because of the time limit. Therefore, the proficiency levels could not be directly compared. The proportions were counted by dividing the number of each inaccuracy by the total number of words read / named (possibilities for inaccuracies) for each subgroup separately. Thus, a percentage of potential occurrences was achieved (see also Protopapas et al., 2013). Second, χ^2 tests were performed to see if the proficiency levels yielded any significant differences in the use of different inaccuracies, comparing the number of children who made no errors vs. children with at least one error of a certain type within the proficiency level. Third, the frequencies and distributions of the different inaccuracy types were counted for each grade, especially to compare the performance in the L2 tasks to the L1 versions *within* the grades. This was considered to reveal possible tendencies between the more and less proficient languages. The control measures, i.e., L1 fluency, to which the L2 results were compared to, are presented in the tables first, followed by the L2 measures.

3.1 Differences in the inaccuracies between the grades

To examine the differences between proficiency levels as regarding the grades, the proportions of inaccuracies in each proficiency level relative to the total number of occasions (words or items) were counted. The largest category of inaccuracies for all groups was the pronunciation problems, and it is presented in a separate figure (Figure 1.) to show the other categories more clearly (in Figures 2.–5.). In RAS L1, there were no differences between the groups in pronunciation difficulties. In the RAS L2, however, Grade 8 results were highly affected by the difficulty in pronouncing the letters A [eɪ] and E [i:] (49% of all their pronunciation problems). The pronunciations for the Finnish vowels are A [ɑ], E [e], and I [i], which causes them to be easily confused in English. The matrix in Grade 4 did not include letters, which is why the results of Grade 4 RAS L2 are not directly comparable to the other grades' results, but the amount of pronunciation problems decreased from Grade 8 to 11. In the Word List, the number of pronunciation problems decreased steadily by proficiency in both language, which indicates that the inefficiency of lexical access decreased with developing skills.

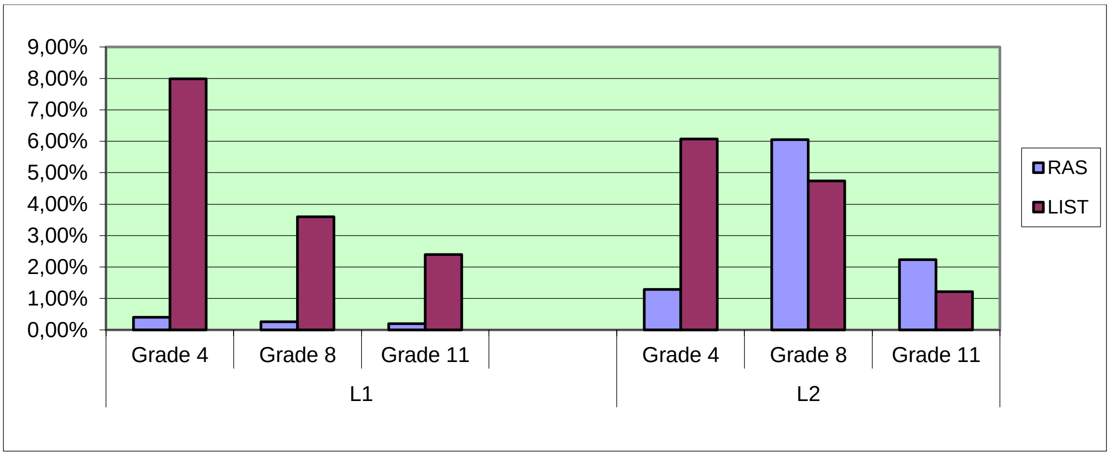

Figure 1. Distributions of relative proportion of pronunciation problems (total number of inaccuracies by category divided by total number of words) by Grades in RAS and Word List tasks.

For the other inaccuracies, in the L1 RAS task, the numbers in both inefficiency and attention-control categories decreased with proficiency (Figure 2.); i.e., most types of inaccuracies decreased with automatization of word retrieval. More skipped and repeated items were found for Grade 4 than for the other grades. In the same-category inaccuracies (SCAT) the groups were indistinguishable. Overall, in RAS L1, even for Grade 4, the accuracy was very high and differences between proficiency levels were quite minimal.

In L2 RAS, the trends were not as clear, and the results of Grade 4 are not directly comparable to the other results. However, the inaccuracies in the inefficiency categories, skipping (SKIP) and same-category errors (SCAT), seemed to decrease with proficiency (Figure 3.). On the other hand, for the attention-control categories, the self-corrections followed an inverted U-shape, as in Grade 8 these were found more often than in the other groups. In repetitions, there were almost no differences between the groups. Guessing, inhibition difficulties, and shifting were almost non-existent in L2.

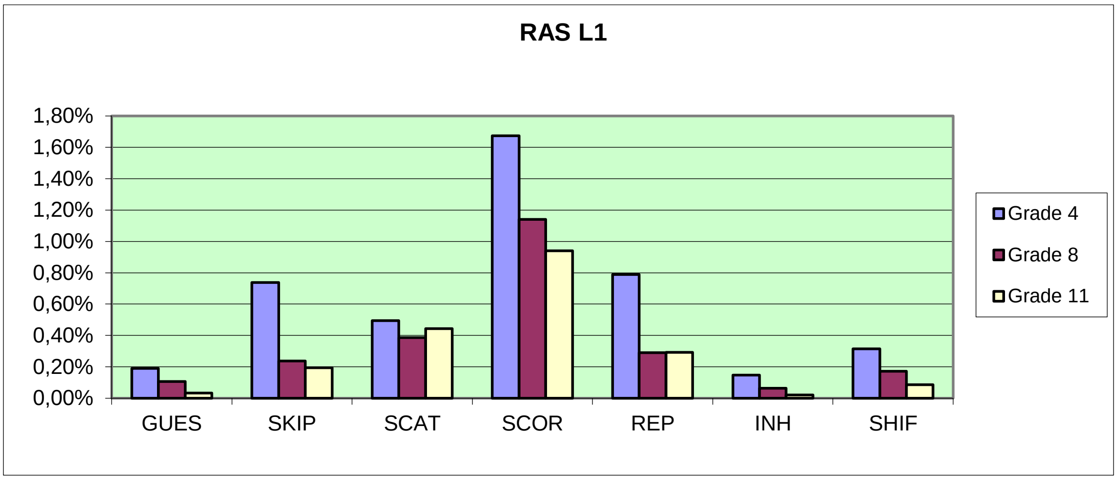

Figure 2. Distributions of relative proportions (the total number of inaccuracies by category divided by total number of words) by Grade in L1 RAS task.

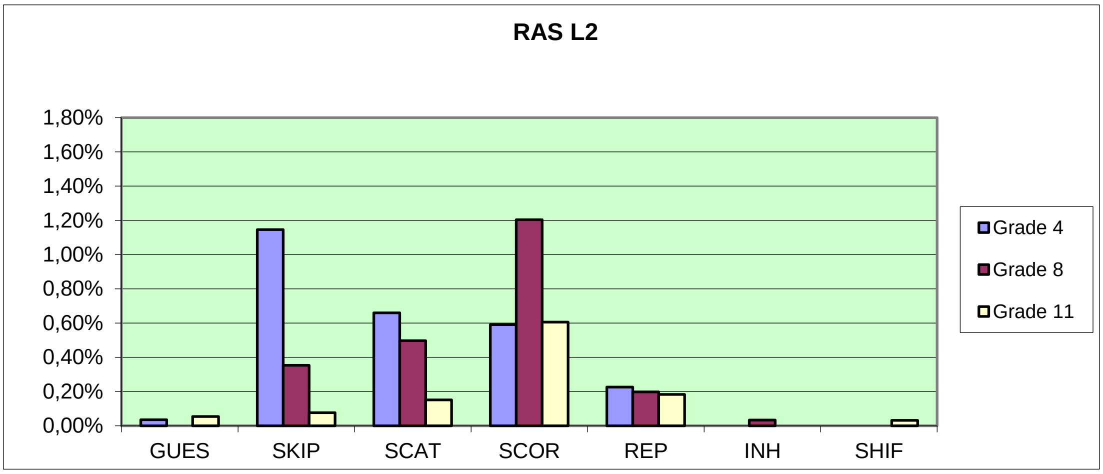

Figure 3. Distributions of relative proportions (the total number of inaccuracies by category divided by total number of words) by Grade in L2 RAS.
Note. GUES = Guessing. SKIP = Skipping. SCAT = Same-category. SCOR = Self-corrections. REP = Repetition. INH = Inhibition. SHIF = Category shifting.

In the Word List task, the inefficient processing in the form of guessing (GUES) decreased by proficiency in both L1 and L2 (Figures 4. and 5.). Skipping was found almost exclusively in L2 and its use decreased by proficiency, but it was used very rarely. For the attentional processes, the number of self-corrections followed slightly an inverted U-shape as in the RAS task: increasing from Grade 4 to Grade 8, and then decreasing again. For repetitions, there were no differences between the grades in L1, and in L2 their number decreased. The number of inhibition difficulties decreased by proficiency and were found almost exclusively in the L1 task.

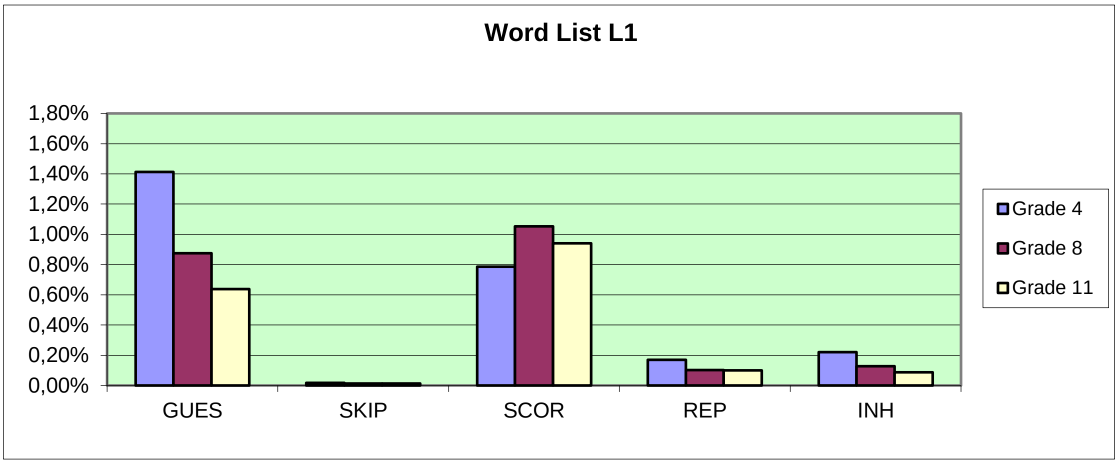

Figure 4. Distributions of relative proportions (the total number of inaccuracies by category divided by the total number of words) by Grade in L1 Word List.

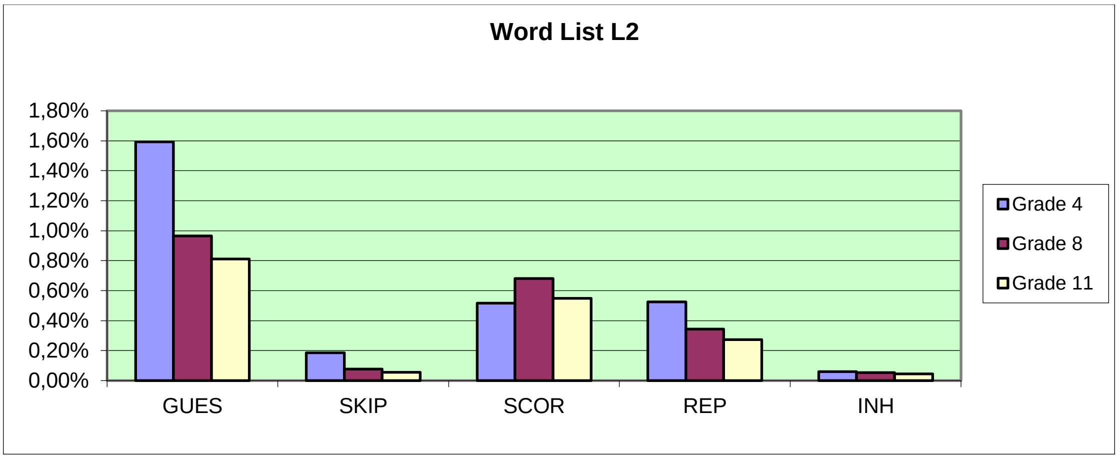

Figure 5. Distributions of relative proportions (the total number of inaccuracies by category divided by the total number of words) by Grade in L2 Word List.
Note. GUES = Guessing. SKIP = Skipping. SCOR = Self-corrections. REP = Repetition. INH = Inhibition.

Second, to test if any of the apparent differences between the grades were statistically significant, χ^2 scores were counted for each inaccuracy category. These compared the proficiency levels in relation to the proportions of children in each grade who used a particular type of inaccuracy. In RAS L2, Grade 4 was left out of the analysis due to the task differences. Grade 8 had more inaccuracies than expected, compared to Grade 11 performance, in self-corrections (χ^2 (1, 427) = 8.44, $p < .01$, adjusted standardized residual = 2.9), in same-category errors (χ^2 (1, 427) = 10.76, $p < .001$, adjusted standardized residual = 3.3), and in pronunciation problems (χ^2 (1, 427) = 8.45, $p < .01$, adjusted standardized residual = 2.9). Other inaccuracy types did not yield significant effects. To compare to RAS L1 results, χ^2 tests showed modestly significant differences between the proficiency levels for self-corrections and repetitions. Grade 4 children differed from the other two grades, as they corrected themselves more than expected (χ^2 (2, 638) = 8.03, $p < .05$, adjusted standardized residual = 2.8) and repeated more items (χ^2 (2, 638) = 6.93, $p < .05$, adjusted standardized residual = 2.5).

In the Word List, the χ^2 test showed significant effects between the groups for self-corrections. There were fewer errors than expected for Grade 4 in both L1 (χ^2 (2, 638) = 22.09, $p < .001$, adjusted standardized residual = -4.6), and L2 (χ^2 (2, 638) = 17.03, p < .001, adjusted standardized residual = -4.1). Furthermore, a significant effect was found for Grade 4 in L2 pronunciation problems, for more than one inaccuracy (χ^2 (4, 638) = 14.47, $p < .01$, adjusted standardized residual = 2.4). To compare to L1 results, more pronunciation problems were found for Grade 4 (χ^2 (2, 638) = 25.22, $p < .001$, adjusted standardized residual = 4.5), and Grade 11 skipped significantly more in L1 than expected (χ^2 (2, 638) = 7.28, $p < .05$, adjusted standardized residual = 2.7). Other inaccuracy types did not yield significant effects.

One additional note is that the patterns from χ^2 tests showed some support for the proportional analyses (see Chapter 3.1) in that all the groups seemed to self-correct less in L2 than in L1. In Word List, self-correcting in L2 was used in Grade 4 by 31% of the children, in Grade 8 by 49%, and in Grade 11 by 49%, as compared to L1 (4: 39%, 8: 57%, and 11: 61%). Similar pattern was found for inhibition difficulties (4: L2 12% vs. L1 19%, 8: L2 15% vs. L1 19%, and 11: L2 21% vs. L1 23%). In RAS, Grades 8 and 11 had fewer difficulties with set-shifting in L2 than in L1 (13% vs. 17% and 16% vs. 19%). In Grade 11, a smaller proportion of students corrected themselves in L2 than in L1 (36% vs. 45%), as well as skipped items (18% vs. 23%).

3.2 Differences in the inaccuracies between the L1 and L2

The frequencies and percentages of different types of inaccuracies across the grades are presented in Tables 2 and 3. The comparisons offer information especially on the group profiles in their performance in L2 as compared to L1, regarding the proportions of the two main categories, i.e., inefficiency-related inaccuracies and the attention-control inaccuracies. Overall, the results were quite similar between the groups regarding the most common types of inaccuracies, especially for the Grades 8 and 11. In RAS L1, the self-corrections (SCOR) constituted the largest proportion of inaccuracies in all grades (4: 35%, 8: 43%, and 11: 43%; see Table 2.). However, Grade 4 had the repetitions (REP, 17%) as the second most common type of inaccuracy, whereas the Grades 8 and 11 had same-category inaccuracies (SCAT, 15%). In RAS L2, the largest group of inaccuracies was the miscellaneous pronunciation problems (PRON) for all groups (4: 33%, 8:

73%, and 11: 67%). The second most common type of inaccuracy in RAS L2 was skipping (SKIP) for Grade 4 and self-corrections (SCOR) for Grades 8 and 11. To summarise, for all the groups, the inaccuracies in L1 belonged mostly to the attention-related categories (SCOR, REP), whereas in L2, the inaccuracies related mostly to inefficiency (PRON, SKIP). However, for the more proficient students, attention-related inaccuracies were the second common type also in L2 (SCOR). A further point to notice is that the last two types of attention-related inaccuracies (inhibition difficulties and shifting), that were rare even in L1, were practically non-existent in L2.

Table 2. Frequencies and percentages of different inaccuracies by language and grade in RAS.

	L1						L2					
Grade	**4**		**8**		**11**		**4**		**8**		**11**	
Inefficiency	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
GUES	18	4.0	10	4.0	3	1.5	2	0.9	0	0	5	1.6
SKIP	70	15.5	22	8.9	18	8.8	66	**28.9**	32	4.2	7	2.3
SCAT	47	10.4	36	**14.6**	41	**20.1**	38	16.7	45	6.0	14	4.5
PRON	38	8.4	24	9.7	18	8.8	74	**32.5**	548	**72.6**	207	**66.8**
Total %		38.3		37.2		39.2		79.0		82.8		75.2
Attention	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
SCOR	159	**35.3**	106	**42.9**	87	**42.6**	34	14.9	109	**14.4**	56	**18.1**
REP	75	**16.6**	27	10.9	27	13.2	13	5.7	18	2.4	17	5.5
INH	14	3.1	6	2.4	2	1.0	0	0	3	0.4	1	0.3
SHIF	30	6.7	16	6.5	8	3.9	1	0.4	0	0	3	1.0
Total %		61.7		62.7		60.7		21.0		17.2		24.9

Note. GUES = Guessing. SKIP = Skipping. SCAT = Same-category. PRON = Pronunciation problems. SCOR = Self-corrections. REP = Repetition. INH = Inhibition. SHIF = Category shifting. Total number of items in the task: L1 Grade 4: 9,500, 8: 9,250, 11: 5,760; L2 Grade 4: 9,300, 8: 9,050, 11: 9,250.

For the Word List, the pronunciation problems constituted the largest percentage of inaccuracies in both L1 (4: 75%, 8: 62%, and 11: 57%) and L2 (68%, 69%, and 41%, respectively; see Table 3.). The most difficult words in L1 Finnish were either the very long ones (*nimikkoluokka* 'dedicated class'), or the ones containing combinations such as diphthongs *kiulu* 'pail' > *kuilu* 'gorge' or liquid consonants *broileri* 'broiler' > *bloireri*. In L2 English, the problematic words often included phonemes that are not part of Finnish phonotactics, such as [ʧ] or [ð]: much > [mɑks], church > [kurs]. The second most common type of inaccuracy in the Word List for Grade 4 was guessing (GUES) in both L1 (13%) and L2 (18%), as well as in

L2 in Grades 8 (14%) and 11 (28%). In the L1 Word List, the Grades 8 and 11 used self-corrections considerably (18% and 24%, respectively). To summarise, the largest proportion of inaccuracies for all proficiency levels and in both languages resulted from inefficiency. It is notable, however, that the proportion of inaccuracies relating to attentional processes increased by grade in both L1 and L2.

Table 3. Frequencies and percentages of different inaccuracies by Grade and language in Word List.

	L1						L2					
Variable	**4**		**8**		**11**		**4**		**8**		**11**	
Inefficiency	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
GUES	167	13.3	137	15.2	103	15.3	188	17.8	163	14.1	145	27.5
SKIP	2	0.2	2	0.2	2	0.3	22	2.1	13	1.1	10	1.9
PRON	944	75.4	564	62.4	387	57.4	717	67.8	801	69.1	217	41.2
Total %		88.9		77.8		73.0		87.7		84.3		70.6
Attention	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
SCOR	93	7.4	165	18.3	152	22.6	61	5.8	115	9.9	98	18.6
REP	20	1.6	16	1.8	16	2.4	62	5.9	58	5.0	49	9.3
INH	26	2.1	20	2.2	14	2.1	7	0.7	9	0.8	8	1.5
Total %		11.1		22.3		27.1		12.4		15.7		29.4

Note. GUES = Guessing. SKIP = Skipping. PRON = Pronunciation problems. SCOR = Self-corrections. REP = Repetition. INH = Inhibition. Total number of items in the task: Grade 4 L1: 11,828, 8 L1: 15,675, 11 L1: 16,148; Grade 4 L2: 11,809, 8 L2: 16,902, 11 L2: 17,865.

4 Discussion

The present study aimed to see whether the differences in the types of inaccuracies in lexical access could be attributed to the L2 proficiency levels. The inaccuracies in lexical access were proposed to stem from two sources: inefficient processing (cf., Walczyk et al., 2004; Grabe, 2009, p. 292) and attention-control (Kormos, 1999, 2000; Gilabert, 2007; Miyake et al., 2000; Segalowitz, 2000, 2010; Jones et al., 2008). It was proposed that these two sources of inaccuracies would behave differently in relation to proficiency: if lexical access is not automatized, there are no resources for attention-control. Therefore, the first category was proposed to decrease with proficiency, whereas the second was hypothesised to be more stable or to increase. For measuring lexical access, one word retrieval (RAS) and one word recognition (Word List reading) task were used. Finnish children learning English from Grade 4, Grade 8, and Grade 11 participated in the study. These grade levels were used as estimates of L2 proficiency. Overall, the findings suggest that the types of inaccuracies were connected to proficiency to some extent: in the L1 tasks and for the older students, the inaccuracies relating to attention-control accounted for the largest proportion of the inaccuracies (self-corrections). However, in L2 and for the Grade 4, the difficulties seemed to stem more from the lack of efficiency (guessing, skipping, and pronunciation problems). Statistically significant differences were found especially for self-corrections (attention-control) and pronunciation problems (inefficiency). These inaccuracies, therefore, were shown to be the most robust indicators of their respective categories. The other categories mostly offered interesting possibilities for future studies, as the number of occurrences was quite limited and the tasks were not designed to measure these categories of inaccuracies.

4.1 Inefficient processing

The inaccuracies connected with the inefficient processing constituted the largest proportion of inaccuracies in the L2 tasks, especially for the youngest group (Grade 4). The number of inefficiency inaccuracies, as well their proportion of total of inaccuracies, was found to decrease with proficiency. These results are in line with the original hypothesis that these types of inaccuracies are connected with lower language proficiency. A few points on the sub-categories are worth noting. First, the miscellaneous pronunciation problems were the largest category of inaccuracies in the L2 tasks as well as in L1 Word List, and proficiency level significantly affected their amount in both languages. Second, guessing in Word List mostly resulted in real words: either conjugated or in some other way modified forms, as in the study by Danielsson (2003). Completely different words were mostly more familiar or frequent items (cf., Ellis, 2002); even the neologisms graphemically resembled the target words. In the RAS, guessing based on form happened very rarely, which is in accord with, for instance, Gollan et al. (2011). Naming the numbers and colours was so automatic that guessing was rarely needed. Comparing the relative proportions, guessing in RAS was more frequent in L1 than L2, and this could indicate too fast a performance, with no time to notice the errors or to spare for correction.

Regarding skipping items, when looking at the relative proportions, in the RAS task children in Grade 4 and 11 seem to have skipped more items in L1 than in L2. In the other instances, in Grade 8 and the Word List task, skipping was found more in L2. These seemingly contradictory results may be due to different reasons for skipping. In the easier tasks or for the more proficient students, skipping may have resulted from the (too) fast performance, but in the more difficult task and for the beginning language learners, there may have been more strategic reasons for skipping difficult-looking items. More support for this interpretation comes from the χ^2 results, which showed that Grade 11 skipped significantly more than expected in the L1 Word List. In the future studies, this interpretation could be examined, for example, with a stimulated recall setting (Kahng, 2014), or eye-tracking paradigm. However, the findings of Marian et al. (2013) of strategic skipping in a Stroop task, offer this hypothesis some confirmation. Furthermore, skipping, when it happened in L1, was often noticed and corrected; in L2, this was not the case. This is in line with the overall processing limitations hypothesis with more resources available in L1 to monitor errors and correct one's own speech (e.g., Kormos, 2000).

4.2 Attentional processes

The inaccuracies connected with the control of attention were found to constitute the largest part of the inaccuracies in the RAS L1 task, and their proportion of the total number of inaccuracies increased with proficiency. Furthermore, they did not follow a decreasing trend, which is in accord with the initial hypothesis that attention-control requires more resources (cf., Lennon, 1990; Gilabert, 2007; Segalowitz, 2010). In the Word List task, Grade 4 self-corrected significantly less than expected in both L1 and L2, as compared to the behaviour of the more proficient students, which further offers support the monitoring resources hypothesis. In L2, the relative proportions of self-corrections increased from Grade 4 to Grade 8 in both tasks. This could indicate that in the beginning of language learning, there is limited capacity and skills to correct one’s speech, and these develop only gradually. On the other hand, it has to be taken into account that the beginning language learners may also lack the L2 skills to correct themselves and also the whole monitoring system is still developing (see e.g., Diamond, 2013). However, in Grade 11 self-correcting decreased again, which may relate to one’s increasing proficiency, when the need for self-correction is smaller.

For the repetitions, there were no clear differences between the groups or languages, which is in accord with the previous findings that they seem to be more related to personal style and strategic use (Engelhardt et al., 2010; Bosker et al., 2013; Peltonen & Lintunen, 2016). The only significant difference was found for RAS L1, where the children in Grade 4 repeated more than the other groups. This could be explained by the strategic use for stalling (cf., Engelhardt et al., 2010). As the Grade 4 children found it often difficult to concentrate on the tasks, it would be very interesting to address if this had any effect on the performance by comparing the distribution of inaccuracies for example along the task. This was, however, outside the current paper's scope.

The tasks used in the present study did elicit only modest amounts of inaccuracies relating to difficulties in inhibiting and set-shifting, and the hypothesised u-shape was not found. There were some tendencies, nevertheless. The number of both difficulties in inhibiting and set-shifting decreased with proficiency, and they were found almost solely in L1, in accordance with previous studies (cf., Favreau & Segalowitz, 1983; Kaukonen & Lanu, 2005; Jones et al., 2008; Marian et al., 2013). These results can be interpreted in line with the hypothesis of easier suppression of irrelevant material and lesser activation of different RAS categories in L2. This would, however, need further confirmation from tasks better designed to measure these processes (see e.g., Diamond, 2013). Nevertheless, the findings concerning attentional processes suggest that these categories do not seem very reliable indicators of proficiency (in line with, e.g., Freed, 2000; Bosker et al., 2013). Therefore, the results of the study call for caution in some of the aspects of the oral fluency assessment; e.g., self-corrections might be better interpreted as active control of the output instead of breakdowns of speech.

4.3 Theoretical implications and limitations of the study

To sum up, the current study assigned the sources of inaccuracies to processing limitations in language use (also Lennon, 2000). The limited language skills can manifest in inefficient lexical access, and when they gradually automatize, resources are freed up for the attention-requiring processes. The current finding that the more proficient student groups used self-corrections more than the less-advanced group is an example of this (also e.g., Lennon, 1990), and in some instances may relate to the strategies of the language users (Kormos, 1999). Thus, the overall proposition is that not only the frequency of the inaccuracies, but also their origin is important when studying the relationship of fluency and proficiency. The inaccuracies originating from the attentional processes seemed more related to higher proficiency, and the automatization of language processes may also lead to involuntary activation of irrelevant material (similarly in the Stroop task: Marian et al., 2013).

The results presented here, although exploratory in nature, nevertheless offer insight into the construct of fluency as a multifaceted phenomenon that should not be treated in a straightforward manner (see also, Kormos, 1999, 2000; Bosker et al., 2013; Peltonen & Lintunen, 2016). There are different trends influencing fluency in speech, especially in L2, and this study highlighted the cognitive processing aspects. One further notable point in the current data is that although L1 data was gathered more as baseline data, lexical access even in L1 was found to be far from fluent. This was especially visible in the Word List task, which also illustrated the different challenges the readers encounter in Finnish and in English. Although Finnish orthography is shallow, the long, inflected words proved to be quite difficult especially under the time-pressure of the current tasks. Therefore, the overall accuracy rates in English were in fact higher in Grade 8 and Grade 11, although a more lenient attitude of the raters in assessing L2 speech has to be considered as well. Overall, these results highlight the difficulty of using strict fluency criteria in assessment. Perceived or global fluency may be a good predictor of L2 proficiency, but when deconstructed, the connections of different phenomena are very hard to pin down (cf., Koponen & Riggenbach, 2000; Freed, 2000; Segalowitz, 2010; Bosker et al., 2013). The framework presented in the current paper is of course only one possibility and other highly interesting aspects for further study include, for example, the emotional issues such as anxiety and its influence on word retrieval and types of inaccuracies (see e.g., Housen & Kuiken, 2009).

For the limitations of the current study, first, it has to be kept in mind that these results apply only to lexical access measurement and not to longer monologue or dialogue settings. On the other hand, following Segalowitz (2010, p. 75–76; also Gholamain & Geva, 1999), lexical access was considered to offer a more pure insight into cognitive fluency, as more general planning and conversational problems were kept to a minimum. Second limitation relates to the tasks used here, as they were not designed to reveal all the types of inaccuracies studied here, and have previously been used for clinical diagnosis of dyslexia. Thus, their purpose originally has not been to distinguish normally developing children in detail, but to diagnose problems. Therefore, the tasks were often not challenging enough for the higher grades to produce enough inaccuracies for the statistical analyses, as the material was much skewed. Nonetheless, these results offer interesting tendencies and questions. Answers to them should be further explored with the help of more fine-tuned and better-suited tasks. More online-methods of measuring would also offer more detailed information on the types of inaccuracies that were only hypothesised indirectly. For instance, skipping might be studied with eye-tracking paradigm to reveal possible differences in the voluntary vs. accidental skipping, and the inhibition difficulties and involuntary activation in lexical access could be tapped with, e.g., bilingual Stroop task.

Endnote

[1] Leech, G., Rayson, R., & Wilson, A. (n.d.) Word Frequencies in Written and Spoken English: based on the British National Corpus. Retrieved October 15, 2010, from http://ucrel.lancs.ac.uk/bncfreq/lists/1_2_all_freq.txt

Acknowledgements

I wish to thank Jaana Haapasalo for her help in the additional data coding, and Lea Nieminen, Kenneth Eklund, Rob Schoonen, and Paula Salmi for their comments on the previous versions of the manuscript. Furthermore, I am grateful to the two anonymous reviewers for their input and insight.

References

Ahonen. T., Tuovinen, S., & Leppäsaari, T. (2003). Nopean sarjallisen nimeämisen testi [The test of rapid serial naming. Research reports.]. Jyväskylä, Finland: Haukkaranta school & Niilo Mäki Institute.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, *29*(2), 175–193.

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*(4), 621–635.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (A travel update). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.) (pp. 285–375). NY: Elsevier.

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159–175.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, *74*(1), 1–15.

Cossu, G., Shankweiler, D., Liberman, I. Y., & Gugliotta, M. (1995). Visual and phonological determinants of misreadings in a transparent orthography. *Reading and Writing: An Interdisciplinary Journal*, *7*(3), 237–256.

Council of Europe (2007). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Danielsson, K. (2003). *Beginners read aloud. High versus low linguistic levels in Swedish beginners' oral reading*. Stockholm, Sweden: Almqvist & Wiksell International.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223–243.

Denckla, M. B., & Cutting, L. E. (1999). History and significance of Rapid Automatized Naming. *Annals of Dyslexia*, *49*, 29-42.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168.

Di Filippo, G., Brizzolara, D., Chilosi, A., De Luca, M., Judica, A., Pecini, C., Spinelli, D., & Zoccolotti, P. (2005). Rapid naming, not cancellation speed or articulation rate, predicts reading in an orthographically regular language (Italian). *Child Neuropsychology*, *11*(4), 349–361.

Ehri, L. (1991). Development of the ability to read words. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (pp. 383–417). NY: Longman.

Eklund, K., Torppa, M., Aro, M., Leppänen, P. H. T., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational Psychology*, *107*(1), 126–140.

Ellis, N. C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Engelhardt P. E., Corley, M., Nigg, J. T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Journal Memory & Cognition*, *38*(5), 617–628.

Favreau, M., & Segalowitz, N. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory & Cognition*, *11*(6), 565–574.

Freed, B. F. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). Ann Arbor, MI: University of Michigan.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science*, *4*(5), 294–298.

Geva, E., & Siegel, L. S. (2000). The role of orthography and cognitive factors in the concurrent development of basic reading skills in bilingual children. *Reading and Writing: An Interdisciplinary Journal*, *12*(1), 1–30.

Gholamain, M., & Geva, E. (1999). Orthographic and cognitive factors in the concurrent development of basic reading skills in English and Persian. *Language Learning*, *49*(2), 183–217.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *45*(3), 215–240.

Gollan, T. H., Slattery, T. J., Goldenberg, D., van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186–209.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461–473.

Häyrinen, T., Serenius-Sirve, S., & Korkman, M. (1999). *Lukilasse. Lukemisen, kirjoittamisen ja laskemisen seulontatestistö peruskoulun ala-asteen luokille 1–6*. [The Lukilasse graded achievement package for comprehensive school age children.] Helsinki: Psykologien Kustannus Oy.

Jones, M. W., Obregón, M., Kelly, M. L., & Branigan, H. P. (2008). Elucidating the componen processes involved in dyslexic and non-dyslexic reading fluency: An eye-tracking study. *Cognition*, *109*(3), 389–407.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854.

Kame'enui, E., & Simmons, D. (2001). Introduction to this special issue: The DNA of reading fluency. *Scientific Studies of Reading*, *5*(3), 203–210.

Kaukonen, S., & Lanu, A.-L. (2005). *Lukemisen osatekijöitä ja ongelmia*. [Components and problems of reading.] Unpublished master's thesis. University of Jyväskylä, Finnish studies.

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, *95*(3), 453–464.

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–22) Ann Arbor, MI: University of Michigan.

Kormos, J. (1999). The effect of the speaker variables on the self-correction behaviour of L2 learners. *System*, *27*(2), 207–221.

Kormos, J. (2000). The role of attention in monitoring second language speech production. *Language Learning*, *50*(2), 343–384.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Ann Arbor, MI: University of Michigan.

Lyytinen, H., Ahonen, T., Eklund, K., Guttorm, T., Kulju, P., Laakso, M.-L., Leiwo, M., Leppänen, P., Lyytinen, P., Poikkeus, A.-M., Richardson, U., Torppa, M., & Viholainen, H. (2004). Early development of children at familial risk for dyslexia – Follow-up from birth to school age. *Dyslexia*, *10*(2), 146–178.

MacLeod, C. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203.

Marian, V., Blumenfeld, H. K., Mizrahi, E., Kania, U., & Cordes, A.-K. (2013). Multilingual Stroop performance: Effects of trilingualism and proficiency on inhibitory control. *International Journal of Multilingualism*, *10*(1), 82–104.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: asymmetrical costs of language selection. *Journal of Memory and Language*, *40*(1), 25–40.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Olkkonen, S., Eklund, K., & Leppänen, P. H. T. (submitted). Fluency of lexical access in L1 and L2 in relation to L2 literacy.

Peltonen, P., & Lintunen, P. (2016). Integrating quantitative and qualitative approaches in L2 fluency analysis: A study of Finnish-speaking and Swedish-speaking learners of English at two school levels. *EuJAL 2016, 4*, 209–238.

Pollatsek, A., Rayner, K., & Balota, D. A. (1986). Inferences about eye movement control from the perceptual span in reading. *Perception & Psychophysics*, *40*(2), 123–130.

Protopapas, A., Fakou, A., Drakopoulou, S., Skaloumbakas, C., & Mouzaki, A. (2013). What do spelling errors tell us? Classification and analysis of spelling errors of Greek schoolchildren with and without dyslexia. *Reading & Writing: An Interdisciplinary Journal*, *26*(5), 615–646.

Riggenbach, H. (Ed.) (2000). *Perspectives on fluency*. Ann Arbor, MI: University of Michigan.

Salmi, P. (2008). *Nimeäminen ja lukemisvaikeus. Kehityksen ja kuntoutuksen näkökulma*. [Naming and dyslexia: Developmental and training perspectives.] Doctoral dissertation. Jyväskylä Studies in Education, Psychology and Social Research, 345.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, *14*, 357–385.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor, MI: University of Michigan Press.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. NY: Routledge.

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: attention-shifting and second language proficiency. *Memory & Cognition*, *33*(4), 644–653.

Simola, J., Holmqvist, K., & Lindgren, M. (2009). Right visual field advantage in parafoveal processing: Evidence from eye-fixation related potentials. *Brain and Language*, *111*(2), 101–113.

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1–14.

Snellings, P., Van Gelderen, A., & De Glopper, K. (2002). Lexical retrieval: An aspect of fluent second language production that can be enhanced. *Language Learning*, *52*(4), 723–754.

Stroop, J. R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, *18*(6), 643–662.

Walczyk, J. J., Marsiglia, C. S., Johns, A. K., & Bryan, K. S. (2004). Children's compensations for poorly automated reading skills. *Discourse Processes, 37*(1), 47–66.

Wolf, M. (1986). Rapid Alternating Stimulus naming in the developmental dyslexias. *Brain and Language*, *27*(2), 360–379.

Received March 18, 2016
Revision received October 7, 2016
Accepted April 6, 2017