Exploring a new technique for comparing bilinguals' L1 and L2 reading speed

Gauvin, H.S.; Hulstijn, J.H.

Published in:
Reading in a Foreign Language

Reading in a Foreign Language April 2010, Volume 22, No. 1
ISSN 1539-0578 pp. 84–103

Exploring a new technique for comparing bilinguals' L1 and L2 reading speed

Hanna S. Gauvin and Jan H. Hulstijn[1]
University of Amsterdam
The Netherlands

Abstract

Is it possible to tell whether bilinguals are able to read simple text in their two languages equally fluently? Is it thus possible to distinguish balanced bilinguals from unbalanced bilinguals with respect to reading fluency in their first language (L1) and second language (L2)? In this study, we avoided making direct comparisons between L1 and L2 reading speeds, comparing, instead, the amount of inhibition caused by a nonlinguistic, external factor (degraded text visibility). In two tasks, 32 university students read 20 target sentences in L1 Dutch and L2 English, each sentence appearing both in normal and in poorly readable font. Degraded font affected reading times substantially, more so in L2 than in L1, as predicted. However, it was not found that participants with higher L2 proficiency were less affected by degraded font in L2 reading than participants with lower L2 proficiency.

Keywords: automaticity, balanced bilingualism, L1 reading, L2 reading, rauding, reading speed, sentence processing

Everyone intuitively understands the loose definition of balanced bilinguals as people equally proficient in both their languages (Schrauf, 2008, p. 114). The question, however, is how to define and measure language proficiency (Hulstijn, in press). As soon as we want to test, in concrete terms, whether someone who is said to be or claims to be a balanced bilingual can indeed perform equally well in both languages, we discover that there do not exist generally accepted valid and reliable tests for this purpose. Degree of bilingualism is normally only assessed with the aid of a profile questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007), not with more objective measures.

In this paper, we focus on testing the ease with which bilinguals read simple text in their two languages. Simple text, in this case, means prose that is easy in terms of linguistic characteristics (composed of short, simple sentences and containing high-frequency words) and has content reasonably easy to comprehend by the reader. This is the kind of reading, without rereading, that Carver (1977) dubbed as *rauding*, a combination of the words reading and auding. "It refers to the frequently occurring situation where individuals are reading or listening, and they are understanding most of the thoughts that they are encountering as they read or aud the sentences involved" (Carver, 1997, p. 6).

There is hardly any empirical research comparing bilinguals' reading speeds in their first language (L1) and second language (L2). Some studies compare the reading processes of highly advanced L2 learners with those of native speaker controls (e.g., Oller & Tullius, 1973), but we found only two studies that compare L2 learners with themselves (i.e., comparing their L2 reading speed with their L1 reading speed). Both studies were conducted by Segalowitz and his associates.

Favreau and Segalowitz (1982) investigated reading and listening speed in both L1 and L2 among two groups of bilinguals in Montreal, Canada. Two English and two French texts with comprehension questions were selected for this investigation. Each participant read one text in English and one in French and listened to one English text and one French text. After reading or listening, they answered comprehension questions. In the listening tasks, participants were presented with time-compressed spoken texts. The maximum speeds at which L1 and L2 texts could be listened to without interfering with comprehension were used in the statistical analyses. In the reading tasks, reading time was clocked with a stopwatch. The order of presentation for language and modality was counterbalanced across participants. Participants were university students, selected for their high proficiency in both English and French, split into a Criterion A group and a Criterion B group (n = 30 in each group, 15 English-French and 15 French-English bilinguals). Students in the Criterion A group reached almost equal levels of reading speed in their L1 and L2 (262 and 275 words per minute [wpm], respectively) and equal speed in listening to L1 and L2 (225 and 224 wpm), while participants in the B Group were much slower in L2 than in L1 in both reading (318 and 234 wpm in L1 and L2, respectively) and listening (250 and 211 wpm in L1 and L2), although reading comprehension scores of both A- and B-group participants were equally high. An interesting additional finding of this study is that the B-group bilinguals read significantly faster in their L1 than the A-group bilinguals (318 vs. 262 wpm), suggesting "the possibility of a trade-off when individuals are highly skilled in two languages" (p. 339). In the A group, 27 out of 30 participants reported having studied in a school where the L2 was the main language of instruction for a period of between 6 and 16 years, while only two students in the B group reported such attendance (6 and 7 years). Thus, it cannot be ruled out that at least some of the A-group students had been exposed to written L2 more than written L1, which might explain their lower L1 reading rates. In a later publication, Segalowitz (1991, p. 62) referred to the A- and B-group participants as *same rate bilinguals* and *different rate bilinguals* respectively. In a later empirical investigation, Segalowitz and Hébert (1990), using the same instruments as Favreau and Segalowitz (1982), replicated the finding that different rate bilinguals read faster in their L1 than same rate bilinguals (313 and 375 wpm).

In Favreau and Segalowitz (1982) and Segalowitz and Hébert (1990), the materials used to assess reading and listening speeds in L1 and L2 were selected from standard text comprehension tests produced by an educational publishing company. According to Favreau and Segalowitz (1982, p. 333), "different texts of the same language had been equated for level of difficulty by the firm that produced them." No information is given how this had been done. While texts between languages were not formally equated, the English and French texts were taken from the highest level of difficulty for their respective language.

It is encouraging to see that these two studies provided some evidence for balanced bilingualism in the same rate readers. However, the evidence is non optimal for two reasons. First, it cannot

be ruled out that participants adopted different strategies in dealing with a potential trade off between reading the text as fast as possible, on the one hand, and memorizing text content in order to answer the text comprehension questions shown after reading (10 and 8 questions in the English and French tasks, respectively), on the other hand. Second, and more importantly, between-language text difficulty was not controlled for. We concur with the authors that, of course, this is "something very difficult to achieve in any event" (Favreau & Segalowitz, 1982, p. 333). Even when translations are used (as in the studies of Bossers, 1991, and Taillefer, 1996, comparing text comprehension, not reading speed, in L1 and L2), sentences in translation pairs may differ in terms of length, morpho-syntactic complexity and lexical frequency. For instance, what may be expressed in one language with a bound morpheme may be expressed in another language with an unbound morpheme, thus compromising the operationalization of reading rate in terms of words read per minute. This leads us to an even more fundamental problem: Languages differ in the linguistic elements they use to express meaning, and thus it is true, in a fundamental sense, that comparing reading rates between languages is a matter of comparing apples and oranges, and so is, a fortiori, comparing text comprehension in L1 with text comprehension in L2.

In a follow-up study, Favreau and Segalowitz (1983) tested the same participants as the ones tested in Favreau and Segalowitz (1982) in an ingeniously designed primed lexical decision task that manipulated expectations about the semantic relatedness of prime and target words and the stimulus onset asynchrony between them. The same-rate bilinguals produced in each language a pattern of reaction times suggesting automatic processing, whereas the different-rate bilinguals did so in L1 but not in L2. Thus, in this study, Favreau and Segalowitz successfully demonstrated that the same-rate bilinguals processed L2 words more automatically than the different-rate bilinguals because they found it harder to inhibit lexical information that was automatically triggered by L2 stimulus words.

Like Favreau and Segalowitz's (1983) follow-up study, our study attempts to circumvent the problem of comparing apples and oranges by *not* making a direct comparison between L1 wpm and L2 wpm but by comparing the degree to which an extra-linguistic task factor impedes L1 and L2 reading. The impeding factor in our study is letter font. On a computer screen, we showed sentences, in both L1 and L2, printed in either clearly or poorly readable font (Times New Roman and Gigi, respectively). For example, the English target sentence number 1 (see Appendix) appeared as I know you did in Times New Roman and, in Gigi, as I know you did.

Participants made plausibility judgments on the contents of these sentences by pressing either a Yes or No key on a computer keyboard. We measured the reaction times (RTs) of the correct responses. The experiment was based on the following idea: For individuals who read L1 and L2 fluently to the same extent, the effect of poorly readable font on plausibility RTs should be equal in L1 and L2. For individuals who read less fluently in L2 than in L1, however, poorly readable font should affect RTs to L2 stimuli more than RTs to L1 stimuli. In other words, we expect an interaction effect of language (L1 vs. L2 stimuli), font (easily vs. poorly readable stimuli) and L2 proficiency. By investigating what poorly readable font does to reading speed in individuals who differ in L2 proficiency, we avoid the problem of comparing RTs of sentences read in L1 and L2 directly. To our knowledge, the use of this technique has never been reported in the published literature. Our study, then, is an exploration of its usefulness.

Method

The design of the study is rather complex. We first give an overview of the study's design and then describe the tasks, materials, and measures in detail.

Two groups of Dutch-L1 university students, differing in English L2 proficiency, performed two computer-administered reading tasks, each task in both L1 Dutch and L2 English. In the story task, participants were given a story to read, sentence by sentence. Some of the story's sentences were marked and participants had to decide as quickly as possible (by pressing one of two keyboard keys) whether the marked sentence fit the story context coherently or not. In the sentence task, participants were shown isolated sentences and had to decide as quickly as possible (by pressing one of two keyboard keys) whether the sentence's meaning was plausible or not. There were 20 target sentences in each combination of task and language. Each target sentence was shown twice, once in normal font (Times New Roman) and once in poorly readable font (Gigi), impeding the ease with which letters and words could be recognized. From here onwards, we refer to this condition as Font and to its levels as "normal" and "degraded." Average reading times of 20 target sentences form the dependent variable in each of the combinations of two languages, two tasks and two fonts. Group (high and low L2 proficiency) formed the between-group independent variable whereas Language (L1 Dutch and L2 English) and Font (normal and degraded font) formed the within-subject independent variables. L2 Vocabulary Size and Short-term Memory Capacity (measured with a digit span task) constituted the two mediating variables (covariates in the statistical analyses). A Latin-square design was applied to control for possible effects of the order in which the tasks in the two languages were administered. The hypothesis to be tested was that the detrimental effect of degraded font in comparison to normal font on the time it took participants to decide whether a target sentence was plausible or not was greater in L2 for the low L2 proficiency group than for the high L2 proficiency group. For L1 this difference was expected not to be found. The hypothesis was tested separately, when sentences were presented in isolation (sentence task) and in story context (story task). We entertained no hypothesis concerning the question of whether the font effect on the low and high proficient groups would be similar or different in the sentence and story tasks.

Participants

Thirty-two students at the University of Amsterdam between 18 and 31 years old and of mixed sex were recruited. Sixteen participants were undergraduate or graduate students in Dutch language and culture (the group of low L2 reading experience, or Low Group). The other 16 participants were enrolled in various English-medium undergraduate or graduate programs (the group of high L2 reading experience, or High Group). We expected the two groups to differ substantially in L2 English proficiency, in particular with respect to reading and vocabulary knowledge, while not differing in short-term memory capacity. These expectations were born out (see the Results section). Of the 32 participants, 5 reported to have a left hand preference; all had normal or corrected-to-normal vision, and were naïve of the purpose of the experiment. Participants signed a consent form before the start of the experiment and received a small fee (€10) for their participation.

Experimental Tasks

Participants performed four computer-administered experimental reading tasks: the L1 and L2 story tasks and the L1 and L2 sentence tasks. In addition, they answered questions concerning their language learning history, self-assessed their English proficiency, and performed an English vocabulary test and a short-term memory test.

Story task: Judging the plausibility of sentences presented in story context. In the Dutch L1 and English L2 story reading task, participants read passages from *The BGF* (short for the "Big Friendly Giant") by Dahl (1982), a book for children, available in both English (the original) and Dutch (a translation). By using the same novel for both languages, we aimed to control for text difficulty level and style of the L1 and L2 reading materials.

Participants read the passages on the computer screen, one sentence at a time. An illustration of the first 45 lines of the English story task is shown in Table 1. When participants pressed the space bar, the current sentence disappeared from the screen and the next sentence appeared. Sentences appeared in black letters (letter size 18) on a silver or yellow background. After participants had read through a sequence of between 2 to 6 story sentences on silver background, the background color turned yellow. This change in background color from silver to yellow signaled participants to decide whether the sentence with the yellow background followed logically in the story so far. They specified their choices by pressing either a green key (yes) or a red key (no) as quickly as possible. After having judged whether the sentence against yellow background fit the story, participants pressed the space bar to proceed. Then, the next story sentence (against a silver background) appeared. After several story sentences, the background color again turned yellow, letting participants know they should make yet another logic decision based on the story's context up to that point. A total of 313 sentences were presented in the L2 story task, 79 of which appeared on yellow background and required a yes-no plausibility response. Of these 79 decision trials, 8 trials required a no-response, while 71 trials required a yes-response. The 8 non-fitting intruder sentences were included only to force participants to pay attention to the contents of the story. We minimized their frequency so as not to disrupt the flow of rauding (Carver, 1977). Of the 71 fitting decision sentences, 40 sentences formed tokens of our target sentences while 31 sentences functioned as fillers. The filler sentences were included in order to make the target sentences less salient as targets in the experiments. There were 20 target sentences, each appearing twice in the story, once in normal font (Times New Roman) and once in degraded font (Gigi). In the case of 10 target sentences, the normal font exemplar appeared before the degraded one; of the other 10 target sentences, the degraded exemplar appeared first. Never did the two members of a target-sentence pair appear in two decision trials in succession. We had manipulated the original text in such a way that the 20 target sentences could naturally appear twice while still fitting the context. Thus, all tokens ($N = 40$) of the 20 target sentences required a yes-response. Because we aimed to investigate fluency in the reading process, we selected target sentences that contained only high-frequency words and were of low grammatical complexity, giving minimal cause to disfluencies caused by lack of lexical or grammatical knowledge. The L1 and L2 target sentences are listed in the Appendix.

In order not to make the 20 target sentences in degraded font (Gigi) too salient, we made all other decision sentences appear in degraded font, too, some in Gigi and others in Mistral. Of the

8 intruder sentences, half appeared in Gigi and half appeared in Mistral. Of the 31 filler sentences requiring a *yes* response, the first one appeared in Gigi as the first decision sentence, to make participants familiar with Gigi before the first target sentence in Gigi appeared, while the remaining 30 sentences appeared in Mistral (see Table 1 for an illustration).

Table 1. *Lines 1-44 of the L2 English story task, by background color, sentence type and font*[a]

	Sentence	Background[b]	Sentence type	Font[c]
1	When Sophie had heard about the giants eating children			
2	she knew they had to do something.			
3	The idea of other children being eaten while she was here with the BFG			
4	had really upset her.			
5	It didn't seem fair.			
6	Potatoes should be cooked for about twenty minutes.	yellow	intruder	Mistral
7	So then she had started thinking.			
8	She thought for a long time.			
9	And then she had it.			
10	She had made a plan.			
11	A plan to rescue the other children from the horrible giants.	yellow	filler	Mistral
12	At first the giant didn't like her plan at all.			
13	He said it was perfectly natural for giants to eats humans.			
14	Even though he didn't like eating them himself.			
15	After a while she had been able to convince the giant.	yellow	filler	Mistral
16	So now they were on their way.			
17	Sophie felt really exited about this.			
18	She had always liked to go traveling and doing good things.			
19	And now she was.	yellow	filler	Mistral
20	The great yellow wasteland lay dim and milky in the moonlight			
21	as the Big Friendly Giant went galloping across it.			
22	Sophie, still wearing only her nightie,			
23	was reclining comfortably in a crevice of the BGF's right ear.			
24	She felt safe now.	yellow	target	Gigi
25	She was actually in the outer rim of the ear, near the top,			
26	where the edge of the ear folds over.			
27	Which under normal circumstances would be a very weird place to be.	yellow	filler	Mistral
28	This folding over bit made a sort of roof for her			
29	and gave her wonderful protection against the rushing wind.			
30	The skin felt soft and warm.			
31	This surprised her.	yellow	target	Gigi
32	Nobody, she told herself, has ever traveled in greater comfort.			
33	Sophie peeped over the rim of the ear			
34	and watched the desolate landscape of Giant Country go whizzing by.			
35	They were certainly moving fast.	yellow	filler	Mistral
36	Sophie had not slept for a long time.			
37	It had been hours since she had gone to bed.			
38	She was very tired.	yellow	target	Gigi
39	Normally she would have been sleeping the past few hours.			
40	But since she had met the giant she hadn't slept at all.			
41	Roses are traditionally used for weddings.	yellow	intruder	Gigi
42	She was also warm and comfortable.			
43	The little girl dozed off.			
44	After a tight sleep she woke up again.	yellow	filler	Mistral

Note. [a]See text for explanations. [b]Sentence background color was silver, if not yellow. [c]Font was Times New Roman (clearly readable), if not Gigi or Mistral (poorly readable).

Different chapters of the novel were used for the Dutch L1 and English L2 story tasks in order to avoid having participants read and digest the same content twice. Because it was difficult to produce experimental story texts with 20 target sentences appearing twice as coherently fitting the context, the number of sentence types and the story length differed slightly between the Dutch L1 and English L2 versions (see Table 2 for details). The total story length was 324 and 313 sentences in Dutch and English, respectively. However, in each story, there were 20 target sentences, each appearing once in Times New Roman and once in Gigi, always coherently fitting the context and thus requiring a yes-response. The main purpose of the task was to produce, for each participant, 20 reaction time (RT) pairs (i.e., for each target sentence, an RT in the normal font condition and an RT in the degraded font condition).

Table 2. *Sentence types and numbers in the L1 and L2 story tasks*

Sentence type	Number of sentences in the L1 story	Number of sentences in the L2 story	Required response
• Total number of sentences	324	313	
• Number of non-decision story sentences (in Times new Roman) appearing on silver background	242	234	Press space bar
• Number of decision sentences appearing on yellow background	82	79	Yes or no
• Target sentences in Times New Roman	20	20	Yes
• Target sentences in Gigi	20	20	Yes
• Intruder sentences in Gigi, requiring a no-response	4	4	No
• Intruder sentences in Mistral, requiring a no-response	4	4	No
• Filler sentence in Gigi (first decision trial)	1	1	Yes
• Filler sentences in Mistral	33	30	Yes

Sentence task: Judging the plausibility of sentences presented in isolation. In this task, semantically plausible and implausible sentences were presented one at a time on a computer screen. Examples of normal, plausible sentences are *Sophie looked at the queen* and *There wasn't any sound*. Examples of abnormal, implausible sentences are *He felt himself lemon* and *Houses tend to walk around a lot*. Participants had to decide as quickly as possible whether the meaning of the stimulus sentence was "normal" by pressing either a green key (normal) or a red key (abnormal) on the keyboard. The sentence remained on the computer screen until the yes- or no-key was pressed. The next stimulus sentence appeared 250 milliseconds (ms) after the response was given. There were 70 stimulus sentences for each language: 20 target sentences as in the story task in Times New Roman font along with 20 identical target sentences in Gigi, 25 plausible filler sentences, and 5 non-plausible filler sentences. All target sentences required a yes-response (as in the story task). None of the target sentences was presented immediately after an abnormal sentence to prevent our measurements from being corrupted by spill-over effects. The stimulus sentences in the L1 and L2 sentence tasks were not translations of each other; they

referred to completely different states of affairs. However, because the L1 and L2 target sentences were taken from the same book, the main character's name occurred in several Dutch target sentences (spelled as "Sofie") as well as in several English target sentences (spelled as "Sophie"). As in the story task, we wanted to decrease the salience of the target sentences in Gigi font. Therefore, the sentences appeared in one normal font (Times New Roman) or in one of two degraded fonts (Gigi and Mistral), as specified in Table 3. As in the L1 and L2 story tasks, the purpose of the L1 and L2 sentence tasks was to compare the RTs of the yes-responses to the 20 target sentences in normal font with those in degraded font.

Table 3. *Sentence types and numbers in the L1 and L2 sentence tasks*

Sentence type	Number of sentences in L1 task	Number of sentences in L2 task	Required response
Total number of trials	90	90	Yes or No
Target sentences in Times New Roman	20	20	Yes
Target sentences in Gigi	20	20	Yes
Implausible sentences in Times	2	3	No
Implausible sentences in Gigi	5	3	No
Implausible sentences in Mistral	3	4	No
Plausible sentences in Times	14	13	Yes
Plausible sentences in Gigi	10	8	Yes
Plausible sentences in Mistral	16	19	Yes

Apparatus. All four experimental tasks (the story and sentence tasks in L1 and L2) were created in E-prime (e-studio 2.0.8.22, Psychology Software Tools 1996–2003) and administered on a Dell Lattitude E5500 notebook. The experiment was programmed in such a way that RTs were measured from the moment the stimulus was presented on the computer screen until a specified response key was pressed on the keyboard.

Order of task administration. All 32 participants performed four experimental tasks (i.e., the sentence and story tasks in L1 and L2). To avoid order effects, we created eight administration orders. Participants either performed the two sentence tasks first followed by the two story tasks, or the other way around. Within each task block, the language order was systematically manipulated. This resulted in eight administration orders. Participants were randomly assigned to these administration orders (4 participants per order).

Non-Experimental Tasks

Language history. Participants filled out a questionnaire concerning their language learning history. They were also asked to mention any known linguistic pathologies such as dyslexia or difficulty in reading in general.

Self-assessment of L2 proficiency. For an estimate of their proficiency in L2 English the participants received a self-assessment grid based on the Common European Framework of Reference for Languages (Council of Europe, 2001). Participants rated, on a six-point scale, their skills on the parameters of listening, reading, spoken interaction, written interaction, spoken production, and written production.

Vocabulary size. Participants performed the Vocabulary Levels Test of English created by Schmitt, Schmitt, and Clapham (2001), based on the test devised by Nation (1983, 1990, 2001). This paper-and-pencil test consists of 50 items. Each item consists of two lists. One list presents six words, and to the right another list provides paraphrases of three of the six words, as the following example shows:

1. business
2. clock ___ part of a house
3. horse ___animal with four legs
4. pencil ___something used for writing
5. shoe
6. wall

Participants' task is to match each paraphrase with the correct word. The test maximum score is 150 (i.e., three points for each item). The test words differ in frequency of occurrence. There are 10 items each at the 2,000-, 3,000-, 5,000-, and 10,000-word-frequency levels while 10 items represent academic vocabulary. Performance was scored both as total number of correct responses, regardless of frequency level (maximum = 150), and as the frequency level obtained, as specified by the test authors.

Short-term memory. Since performance in the main experimental tasks (reading sentences and judging the plausibility of sentence meanings as quickly as possible) might be mediated by short-term memory, we included the administration of a backwards computer-administered digit span task in the design of the study.[2] The stimuli, consisting of a series of digits, ranging in length from 2 to 9 digits, were visually presented on the computer screen digit by digit with 1-second intervals. Participants keyed in their responses on the keyboard. There were two trials for each length. The span score was determined by the highest digit-number length for which both trials could be correctly reproduced backwards.

Procedure

Participants were tested individually in an office at the University of Amsterdam.
The session, which lasted between 45 and 75 minutes, comprised of the following tasks. First, participants read and signed a consent form. They then filled out the language history form, completed the L2 self-assessment grid, and performed the vocabulary-size and the digit-span test. Then, the four experimental tasks followed. Task order was systematically manipulated across participants as described above. Participants received instructions for the experiment on the computer screen. In these instructions participants were asked not to change answering strategies during the experiment. This was again stressed orally before the start of the experiment.

Results

In this section we first report on the potential mediating variables, short-term memory capacity, self-reported L2 proficiency, and L2 vocabulary. We then report on the effect of degraded font

on RTs in the experiment examining whether there was evidence for the expected Language × Font × Proficiency interaction in the story and the sentence task (ANOVA approach). We conclude this section with analyses of individual differences in performance in the experimental tasks, taking the data of the two proficiency groups together (correlation approach and examination of individual cases).

Short-Term Memory Capacity

Short-term memory capacity, as measured with the backward digit span task (M = 6.8, SD = 1.1) was not associated with the eight experimental measures (Pearson's r ranged from -.320 to -.45, all coefficients non-significant, with N = 32). Neither was there any association between digit span and scores on the English vocabulary test as scored in terms of frequency level or number of items correct (r = .045 and -.106, respectively). In the remainder of this section, we will therefore not take digit span into account.

L2 Knowledge

We first checked whether students in the Dutch language and culture program (the Low Group) did indeed differ in English L2 skills from the students enrolled in English-medium language programs (the High Group). Using the self assessment grid in the Common European Framework of Reference for Languages (Council of Europe, 2001), both groups gave themselves the same scores for listening. In all other domains, however, participants in the Low Group assessed themselves to be at a lower level than participants in the High Group.

From the scores on the Vocabulary Levels Test we calculated a vocabulary level for every participant, as specified by the test designers (Schmitt, Schmitt, & Clapham, 2001). In the High Group, 14 out of 16 participants obtained the highest level score (10,000 words). The other 2 obtained the second highest level score (5,000 words). In the Low Group, only 3 participants managed to reach the 10,000 word level; 6 attained the 5,000 word level, while all others reached an even lower level. Under the method of scoring the number of correct responses, regardless of frequency level (Max = 150), the Low Group (M = 115, SD = 19, range = [77–145]) and the High Group (M = 143, SD = 8, range = [124–150]) performed significantly differently from each other, $t(30)$ = -.5390, $p < .001$. The Pearson correlation between the level scores and the number of correct responses was .93 ($p < .001$, N = 32).

In conclusion, although the Low and High groups differed in L2 knowledge, there was considerable dispersion in the Low Group and a partial overlap between the groups. We will return to this observation below.

The Experiment

Data cleaning. From each participant we obtained 160 reaction time measures—RT responses to 20 target sentences in each condition (two tasks, two languages, two fonts). From this data file, measures were excluded if incorrect responses were made (i.e., if a sentence had been judged as implausible when it should have been judged as being plausible). The corresponding responses in the other font condition were also deleted. Thus, for example, if we deleted the RT of the

incorrect response to target sentence 4 in task 2 in the normal-font condition, we also deleted the RT of the response to sentence 4 in task 2 in the degraded-font condition. This resulted in a loss of 8.7% of the data, and 4% of this loss can be attributed to a small programming error which caused one target sentence in the sentence task to appear twice in a normal condition instead of once in normal condition and once in degraded condition. An analysis of the incorrect responses reveals that the Low Group made more errors (245) than the High Group (198), which was also reflected in the errors per task. However, a *t* test revealed this difference between groups not to be significant.

The file was also checked for extreme values. Any RT that differed more than 3,000 ms between the two font conditions was deleted from the file. We did not find it credible that, with a grand mean RT of approximately 1,400 ms such a big difference between two conditions on the same target sentence would reflect valid responses. This removal of data together with the previous removal adds up to a total loss of 10% of the data.

On the remaining data, arranged as participant data and as target-sentence data, repeated-measures ANOVAs were conducted, with alpha set at .05. We report the analyses on the participant data; the analyses on the target-sentence data produced a similar pattern of results.

Effects of order. The participants who completed the sentence task before the story task performed somewhat slower, but significantly so, on the sentence task in L1, $t(30) = 3.721$, $p = .001$, and on the sentence task in L2, $t(30) = 2.904$, $p = .007$, than the participants who performed the story task before the sentence task. For the story task, however, no significant order effect was obtained. Thus, it might have been the case that the story task provided a framework that mildly facilitated recognition of the target sentences in the sentence task but the sentence task did not prime the target sentences in the story task. There was no effect of language; whether a task was completed in L1 before L2 or vice versa did not significantly affect target-sentence RTs. Because the eight orders were equally distributed between the two groups, and participants were assigned randomly to these orders, we did not include Order as a factor in the analyses reported in the remainder of this section.

Main effects and interactions. Our hypothesis was that we expected the slow down in RT caused by degraded font to be equal for the Low and High groups in Dutch L1 but that the Low Group would be affected by degraded font in L2 English more so than the High Group. The descriptive statistics are shown in Table 4.

We conducted repeated measures ANOVAs on the RT data in the sentence and story task separately, with Language (L1 vs. L2) and Font (normal vs. degraded) as the two within-subject factors and Group (Low vs. High) as the between-subject factor. First, we obtained a significant main effect of Language in the sentence task, $F(1, 30) = 25.951$, $p < .001$, partial $\eta^2 = .464$, and in the story task, $F(1, 30) = 20.224$, $p < .001$, partial $\eta^2 = .403$. This finding reflects the fact that, on average, it took participants longer to pass their plausibility judgments in L2 than in L1. This does not concern our research question because straight cross-language comparisons are confounded with language and material differences, as we argued in the Introduction.

Large main effects were also found for Font in the sentence task, $F(1, 30) = 62.921$, $p < .001$,

partial $\eta^2 = .677$, and in the story task, $F(1, 30) = 51.733$, $p < .001$, partial $\eta^2 = .633$. This finding shows that we had been successful in manipulating the visibility of the target sentences. In L1, degraded font caused an RT slow down of 141 and 187 ms in the sentence and story task, respectively. In L2 the slow down was 341 and 252 ms in the sentence and story task, respectively. However, the Language × Font interaction, although significant in the sentence task, $F(1, 30) = 23.089$, $p < .001$, partial $\eta^2 = .435$, just missed significance in the story task, $F(1, 30) = 3.901$, $p = .058$, partial $\eta^2 = .115$.

Table 4. *Reaction times (in ms) by task, language, font, and group*

Language & font	Group	*n*	Sentence task		Story task	
			M	*SD*	*M*	*SD*
L1 Normal	Low	16	1,179	325	1,234	219
	High	16	1,340	315	1,111	358
	Total	32	1,259	325	1,172	298
L1 Degraded	Low	16	1,349	288	1,451	273
	High	16	1,452	327	1,266	291
	Total	32	1,400	307	1,359	293
L2 Normal	Low	16	1,532	516	1,481	359
	High	16	1,403	383	1,312	448
	Total	32	1,468	451	1,396	408
L2 Degraded	Low	16	1,880	478	1,794	424
	High	16	1,738	473	1,502	352
	Total	32	1,809	473	1,648	411

The between-subject Group factor, comparing the 16 students in the L1-medium university program (Low Group) with the 16 students in the L2-medium program (High Group), was neither significant in the sentence task nor in the story task. The Language × Group interaction in the sentence task was significant, $F(1, 30) = 4.878$, $p = .035$, but small (partial $\eta^2 = .140$) while not significant in the story task, reflecting the fact that our Low and High participant groups did not sufficiently differ in L2 reading speed. The Language × Font × Group interaction, pertaining to our hypothesis, was neither significant in the sentence task ($p = .59$) nor in the story task ($p = .37$). Thus, although degraded font did slow down RTs in L2 more than they did in L1, this effect was not mediated by the group factor.

Identical ANOVAs on the target-sentence data (so called F1 analyses) produced the same pattern of results. We also split the participants into groups differing in their vocabulary size scores (the Low Group [$n = 15$] with scores ranging from 77 to 129, and the High Group with scores ranging from 138 to 150 [$n = 17$]), but ANOVAs with this between-group factor did not produce, in either the sentence or the story task, the Language × Font × Group interaction that we had hoped to find.

Furthermore, we computed the slow down ratios caused by degraded font by dividing the mean RT for sentences read in degraded font by the mean RT for sentences read in normal font, for each participant, in each language and group combination, in the sentence and story tasks

separately. Table 5 shows the resulting percentages.[3] Although the figures in the sentence task show a remarkably low percentage of delay in L1 RTs for the High L2 Group (9%) and although the sentence task data therefore appear to reflect a significant Language × Group interaction, such an effect was not obtained. Neither was the Language × Group interaction significant in the story task data.

Table 5. *Mean percentages of delay in RTs caused by degraded font, by task, language and group*

Group	Sentence task		Story task	
	L1	L2	L1	L2
Low	17	27	18	22
High	9	25	17	18

In conclusion, although we had been successful in manipulating the font factor, in that it caused an overall slow down in decision times in the two plausibility-decision tasks, and although degraded font slowed down decision times in L2 more than it did in L1 in both the sentence and the story task, this was not the case differentially for the students in the L1- and L2-medium university programs, nor for the students with smaller and larger L2 vocabularies.

A closer look at individual differences. As the analyses reported above did not produce evidence for the hypothesis that participants in the Low Group would be affected by degraded font in L2 more than participants in the High Group, we started to analyze the data of all 32 participants together, in search of individual differences that might be revealing with respect to our research question. We first adopted a correlational approach. The sentence and story task each produced four scores for each participant: mean RTs for L1 normal font, L1 degraded font, L2 normal font, and L2 degraded font. Each mean was computed from RTs on 20 target sentences.

The four highest between-variable correlations were obtained for the four font comparisons: RTs on normal and degraded target sentences correlated substantially: sentence task L1 (r = .80), sentence task L2 (r = .90), story task L1 (r = .84), and story task L2 (r = .85), with $p < .001$ and $N = 32$ in all correlations. Interestingly, cross-task correlations on identical sentences were relatively low: L1 normal font, r = .51, p = .003, L1 degraded font, r = .36, p = .04, L2 normal font, r = .38, p = .034, and L2 degraded font, r = .15, ns), suggesting that deciding on the plausibility of sentences presented in isolation is a process different from deciding on the plausibility of exactly the same sentences when presented in a story context.

In both tasks and both languages, fast readers in the normal condition were less affected by degraded font than slow readers because there was a modest but significant association between mean RT in the normal condition and the proportion of slow down caused by degraded font: for L2, r = -.45, p = .01 (story task) and r = -.51, p = .003 (sentence task); for L1, r = -.43, p = .013 (story task) and r = -.53, p = .002 (sentence task).

We then conducted an exploratory regression analysis with RT in the L2 degraded font condition in the *story task* as the dependent variable, and with RT in the L2 normal, L1 degraded, L1 normal conditions, proficiency group (high vs. low), L2 vocabulary score, and digit span as the independent or predictor variables. All predictor variables were entered simultaneously. The total

variance explained was 86%. The contribution of digit span and vocabulary size was not significant, proficiency group was marginally significant ($p = .045$), while the contribution of the three RT measures was significant ($p < .005$ for each). We then conducted a forced entry regression analysis with RT in the L2 normal condition in the first block and proficiency group in the second. The first step of this analysis, with RT in the L2 normal condition as predictor, explained 72% of the variance, $F_{change}(1, 30) = 77.774$, $p < .001$. The second step, with proficiency group as predictor, moved the total variance explained 4% higher but the effect of this predictor was insignificant ($p = .053$). When we changed the order of entry, entering proficiency group first and RT in the L2 normal condition second, proficiency order explained only 13% of the variance, $F_{change}(1, 30) = 4.463$; $p = .043$, and, of course, RT in the L2 normal condition being highly significant, $F_{change}(1, 30) = 74.397$, $p < .001$. Similar regression analyses on the data of the *sentence task*, with RT in the L2 degraded condition as the dependent variable, produced only one significant predictor, viz., RT in the L2 normal condition, explaining 81% of the variance, $F_{change}(1, 30) = 126.373$, $p < .001$.

It is clear from these results that, while our 32 participants did differ in L2 vocabulary size (the scores on the vocabulary test ranged from 77 to the maximum score of 150, with a mean of 129 and a standard deviation of 20), individual differences in reading speed far outweighed the individual differences in vocabulary size.

Next, we looked at the data of some participants more closely. There were 5 participants who scored 150 (maximum) and 2 who scored 149 on the L2 vocabulary test. Were these 7 individuals, who knew so many L2 words, affected by degraded font in L2 to the same extent as in L1? With one exception, none of these seven participants exhibited a delay in L2, in both tasks, equal to or smaller than in L1. The exception, however, is one individual (Subject 15) with delays of 33% and 30% (story task, L1 and L2, respectively) and 37% and 36% (sentences task, L1 and L2, respectively).

Of the 10 fastest readers in the L2 normal condition in the story task, 5 were also among the 10 fastest readers in the L2 normal condition in the sentence task. Of these 5 individuals, only 2 were delayed by degraded font to an equal extent, or even less so, in L2 in comparison to L1. The first participant, Subject 15, mentioned in the previous paragraph, was affected by degraded font equally in L2 and L1 in both tasks. The other participant was Subject 5, who scored 119 on the L2 vocabulary test. He or she was affected by degraded font even less in L2 than in L1 in each task: delays of 21% and 8% (story task, L1 and L2, respectively) and 45% and 20% (sentence task, L1 and L2, respectively). Should we call these two participants balanced bilinguals then? This remains to be seen because one of the slowest participants (Subject 10), with mean RTs of 2,041 and 2,089 ms, respectively, in the L2 normal font condition in the story and sentence tasks, was also affected by degraded font in L2 less than in L1, with delays of 47% and 15% (story task, L1 and L2, respectively) and 21% and 7% (sentences task, L1 and L2, respectively). However, this participant was the poorest performer of all in the L2 vocabulary test (77 out of 150). Clearly then, Subject 5 only managed to be affected by degraded font in L2 less than in L1 by reading very slowly. On balance then, in our sample of 32 individuals, there was one participant who, on the basis of (a) high L2 vocabulary knowledge, (b) high reading speed in L1 and L2, and (c) equal slow down in L1 and L2 caused by degraded font, might be called a balanced bilingual.

Discussion

Is it possible to tell whether bilinguals are able to read simple text in their two languages equally fluently? Is it thus possible to distinguish truly balanced bilinguals from unbalanced bilinguals with respect to reading fluency in L1 and L2? In the Introduction, we argued that comparing bilinguals' reading speeds, expressed as the number of words read per minute (wpm), constitutes a non-optimal method because of differences between languages in which units count as words and appear as visually separate units in print. Our study attempted to circumvent the problem of comparing wpm in L1 and wpm in L2. Instead, we compared the degree to which an extra-linguistic task factor (poorly readable letter font) impedes L1 and L2 reading. In two tasks (reading of sentences in a story context and reading of sentences in isolation), 32 university students in the Netherlands, read 20 target sentences in L1 Dutch and L2 English, each sentence appearing both in normal font (Times New Roman) and in poorly readable, degraded font (Gigi). Sentences appeared on the computer screen one by one and participants made plausibility judgments with respect to their content by pressing either a yes- or no-key. Mean reaction times (RTs) to correct responses in each condition constituted the dependent variable. We expected that, for individuals who read L1 and L2 fluently to the same extent, the effect of poorly readable font on plausibility RTs should be equal in L1 and L2. For individuals who are less fluent readers in L2 than in L1, however, poorly readable font should affect RTs to L2 stimuli more than RTs to L1 stimuli. Participants were divided into a Low Group, consisting of students in Dutch-medium programs, and a High Group, consisting of students enrolled in English-medium programs. In addition to performing the experimental reading-speed tasks, participants were administered a digit-span task as a measure of short-term memory, and an L2 vocabulary test; they also self-assessed their L2 proficiency.

The findings clearly show that degraded font, in the sample of 32 L1 Dutch users of L2 English, on average, slowed down RTs significantly and substantially, and, as expected, more so in L2 than in L1, both in the story task and in the sentence task. Crucially, as predicted, degraded font affected the processing of the linguistic information, rather than the processing of higher-order information, involved in establishing sentence meaning, judging the sentence's plausibility, or executing the motor response. We believe the font-induced difference in delay to reflect a difference in the level of reading automaticity in the two languages (Favreau & Segalowitz, 1983; LaBerge & Samuels, 1974; Raduege & Schwantes, 1987). On average, our participants were less automatic in processing linguistic information in L2 than in L1. Thus, our new technique proved its value as a research tool.

However, we did not obtain the expected Language × Font × Group interaction. In other words, the Low Group participants were not slowed down in L2 by degraded font more than the High Group participants. We believe that the two groups did not sufficiently differ in L2 proficiency for the three-way interaction to obtain. Performance on the L2 vocabulary test showed a large dispersion of scores and overlapping variances between the two subgroups. Should we have selected two groups of L2 users further apart from each other in terms of L2 proficiency, we might have found the expected interaction. We return to this point shortly.

When we analyzed the RT data of all participants together (i.e., not comparing the two L2 proficiency groups), we observed that reading speed in the degraded font conditions, in both L2 and L1, was first and foremost predicted by reading speed in the corresponding normal font conditions, and that reading speed in the normal font conditions differed widely among participants. This is a remarkable finding, given that all participants were university students (i.e., individuals who read far more and are brighter than the population average, and can therefore be assumed to read faster and make plausibility judgments faster). Thus, even if we had recruited two L2 user groups wide apart from each other in terms of L2 knowledge and skills, an effect of L2 proficiency might still not have obtained because of the overwhelming power of the individual differences in decision times in reading L1 text. The overwhelming effect of individual differences in processing speed underlines the importance of comparing individuals with themselves, performing in L2 and in L1, while downplaying the weight of comparing groups of individuals, differing in L2 declarative knowledge. This applies to L1 reading by monolinguals, too. The fact that person A reads a given text faster than person B, does not have to mean that B processes linguistic information less automatically than A. It may well be that B takes more time for semantic, interpretative processing than A. To exclude semantic processing from the experimental task, so that participants only have to engage in linguistic processing at lower, non-semantic levels, would require giving subjects an experimental task different from the one we used. For instance, one might present subjects with grammatical, but semantically unpredictable sentences, such as *The pain hung near the large tube* or *How does the size learn the fine plane?* In a self-paced reading task, participants read the sentences word-by-word or phrase-by-phrase as fast as possible and reproduce the sentence after reading. Reading times of sentence fragments (of sentences correctly reproduced) would form the dependent variable. Font (normal vs. degraded) and language (L1 vs. L2) could be manipulated as in the experiment reported here.[4]

When we analyzed the RT data case by case, we found that there were three participants who exhibited delays in L2 reading speed equal to delays in L1 reading speed. One of them was a slow reader with poor L2 vocabulary knowledge, one of them read fast but his or her L2 vocabulary knowledge was not at ceiling, and one was relatively fast and performed at ceiling on the L2 vocabulary test. On the basis of our technique and its rationale, only the third participant might then be called a balanced bilingual in terms of reading simple text.

Another interesting finding from this research is that the task of deciding on the plausibility of the target sentences and pressing the appropriate key was performed faster when sentences were presented in the context of a story than when presented in isolation. This is remarkable because the story task not only required pressing one of two keys as in the sentence task, but also switching from pressing the space bar (to obtain the next story sentence on the screen) to pressing a key on the keyboard to submit the plausibility decision. The finding might be explained by a facilitating effect of the context in the story reading task. Reading is not only a bottom-up process but also involves top-down processes allowing readers to make predictive inferences from the text's unfolding story (see, for instance, Calvo, Meseguer, & Carreiras, 2001; Hess, Foss, & Carroll, 1995; Hoeks, Stowe, & Doedens, 2004). The prediction of the upcoming text facilitates the recognition of the words in the next sentence, and can thereby make comprehension of a sentence in the context of an unfolding story to be faster than comprehension of the same sentence without context.

In conclusion, the technique explored in this study has proven its value in that we could show that it is possible to avoid making direct comparisons between the L2 reading speed and the L1 reading speed, while, instead, comparing the amount of slowdown in L1 and L2 reading speeds, caused by a nonlinguistic, external factor (degraded text visibility). In future research, we intend to explore the use of semantically unpredictable sentences in a non-semantic reproduction task, in comparison to the use of meaningful sentences in a plausibility judgment task.

Notes

1. The study reported in this article was conducted by the first author, a graduate student in the Cognitive Science program at the University of Amsterdam, under the supervision of the second author. The second author, Jan Hulstijn, wishes to dedicate this article to Paul Nation, whom he got to know as both a pioneering scholar and a wise person.

2. The digit-span task was created by Nomi Olsthoorn and Sible Andringa, at the Amsterdam Center for Language and Communication of the University of Amsterdam. We would like to thank them for allowing us to use this test.

3. To assess what reading text in degraded font did to reading text in normal font, one can wonder whether it is better to divide the RTs in the degraded font condition by the RTs in the normal condition (which is what we report in the main text) or to subtract the RTs in the degraded condition from those in the normal condition. To remain on the safe side, we computed the proportion scores as well as the subtraction scores and found them to correlate highly with one another (r = .97, .86, .97, and .94 in the L1 sentence, L2 sentence, L1 story, and L2 story tasks, respectively). Note that these high correlations do *not* reflect a mathematical necessity.

4. Engelen and Hulstijn (Engelen, 2009) exposed Dutch university students with syntactic prose in Dutch L1 and English L2 in an aural task. Stimulus sentences were either played in clear speech or under white noise at a signal-to-noise ratio of -3 dB. Participants wrote down what they heard (dictation task). In this experiment, response *accuracy* (number of content words correctly reproduced) was the dependent variable. Processing *speed* was not investigated. White noise had a detrimental effect on performance, more so in L2 than in L1—a finding similar to the effect of degraded font on reading speed in L1 and L2, obtained in the experiment reported here.

References

Bossers, B. (1991). On thresholds, ceilings and short-circuits: The relation between L1 reading, L2 reading and L2 knowledge. *AILA Review*, *8*, 45–60

Calvo, M. G., Meseguer, E., & Carreiras, M. (2001). Inferences about predictable events: Eye movements during reading. *Psychological Research*, *65*, 158–169.

Carver, R. P. (1977). Toward a theory of reading comprehension and rauding. *Reading Research Quarterly*, *13*, 8–63.

Carver, R. P. (1997). *Reading rate: A review of research and theory*. San Diego, CA.: Academic Press.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press / Council of Europe.

Dahl, R. (1982). *The BFG* (the Big Friendly Giant). London: Cape.

Engelen, J. (2009). *First and second language listening ability in noise: A within-subject comparison*. Unpublished master thesis, Faculty of Humanities, University of Amsterdam.

Favreau, M., & Segalowitz, N. S. (1982). Second language reading in fluent bilinguals. *Applied Psycholinguistics*, *3*, 329–341.

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

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*, 62–82.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*, 59–73.

Hulstijn, J. H. (in press). Measuring second language proficiency. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (EMLAR). Amsterdam: Benjamins.

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

Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 1–28.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, *5*, 12–25.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

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

Oller, J. W., Jr., & Tullius, J. (1973). Reading skills of non-native speakers of English. *IRAL*, *11*, 69–80.

Raduege, T. A., & Schwantes, F. M. (1987). Effects of rapid word recognition training on sentence context effects in children. *Journal of Reading Behavior*, *19*, 395–414.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*, 55–88.

Schrauf, R. W. (2008). Bilingualism and aging. In J. Altarriba & R. R. Heredia (Eds), *An introduction to bilingualism: Principles and processes* (p. 105–127). New York: Erlbaum.

Segalowitz, N. (1991). Does advanced skill in a second language reduce automaticity in the first language? *Language Learning*, *41*, 59–83.

Segalowitz, N., & Hébert, M. (1990). Phonological recoding in the first and second language reading of skilled bilinguals. *Language Learning*, *40*, 503–538.

Taillefer, G. F. (1996). L2 reading ability: Further insight into the Short-Circuit Hypothesis. *The Modern Language Journal*, *80*, 461–477.

Appendix A

L2 English Target Sentences

1	I know you did.
2	This surprised her.
3	She looked frightened.
4	Sophie looked at the queen.
5	She was lost.
6	The queen looked at the maid.
7	That is what I dreamt.
8	She felt safe now.
9	There wasn't anybody in the garden.
10	You're awake now.
11	She looked as though she was going to faint.
12	She was very tired.
13	This felt strange for her.
14	It was still dark.
15	There wasn't any sound.
16	She was still sitting in the window-sill.
17	The very idea of what was happening was absurd.
18	There is no such thing as giants.
19	I am sure.
20	Sophie never thought to ever be this close to the queen.

L1 Dutch Target Sentences

1	Ze keek naar de maan.
2	Zonder bril kon ze bijna niets zien.
3	Ze had kippenvel van top tot teen.
4	'Wat stinkt dit' dacht Sofie.
5	Sofie stond op.
6	Zijn voeten waren enorm groot.
7	Hij was zo groot als een huis.
8	Hij pakte Sofie vast.
9	Als ik maar niet wordt opgegeten, dacht Sofie.
10	Het zag er gek uit.
11	Zoiets had ze nog nooit gezien.
12	Ze kon bijna niet geloven wat ze zag.
13	Het zag er onecht uit.
14	In het huis was het volkomen stil.
15	Hij keek Sofie aan.
16	Wat zag hij er eng uit.
17	'Dat lust ik niet' zei hij
18	Haar hele lijf was verstijfd van schrik.
19	Dit kostte hem geen moeite.
20	Hij zette een grote stap.

About the Authors

Hanna S. Gauvin studied linguistics at Leiden University and is currently enrolled in the cognitive science research-master program at the University of Amsterdam.

Jan Hulstijn (PhD, University of Amsterdam, 1982) is a professor of second language acquisition at the University of Amsterdam, Faculty of Humanities, Amsterdam Center for Language and Communication (ACLC). Most of his research is concerned with cognitive aspects of the acquisition and use of a nonnative language (explicit and implicit learning; controlled and automatic processes; components of second-language proficiency). His webpage (http://home.medewerker.uva.nl/j.h.hulstijn/) provides information concerning his research projects and publications. E-mail: J.H.Hulstijn@uva.nl