The effectiveness of comprehensive corrective feedback in second language writing
van Beuningen, C.G.

Citation for published version (APA):
Chapter 5

An in-depth analysis of the effects of direct and indirect corrective feedback in L2 writing

Qualitative study

5.1 Abstract

The value of corrective feedback (CF) for second language (L2) development has been debated ever since Truscott’s (1996) article in which he claimed error correction to be ineffective and potentially harmful. Recent studies have shown, however, that written CF does promote L2 learners’ accuracy development. Nevertheless, little is known about the ways in which individual learners engage with and benefit from the feedback they receive, because most studies assessed accuracy gains quantitatively in terms of global measures (e.g. error rates). The present multiple case-study set out to fill this gap by analyzing the accuracy performance of four L2 learners of Dutch in detail. In doing so, we explored the amenability of different error types (e.g. lexical errors, morphosyntactic errors) to direct and indirect CF. Findings proved that in-depth accuracy analysis is a valuable addition to global measurement of accuracy development, because it can reveal details on CF effectiveness that will not be unveiled by global analyses. Our study showed, for example, that the efficacy of CF may be mediated by factors such as learners' level of successful CF uptake, or the nature of the targeted error.

1 A slightly adapted version of this chapter has been submitted as: Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (under review). Corrective feedback in L2 writing: An in-depth analysis of the effects of direct and indirect corrective feedback in second language writing.
5.2 Introduction

Although common practice in L2 instruction, the value of CF for accuracy development has been a fiercely debated topic (see particularly Ferris, 1999; 2004; Truscott, 1996; 1999; 2007; Truscott & Hsu, 2008). Truscott (1996) even summoned the abandonment of CF from L2 classrooms. He argued that, based on insights from SLA theory, CF would be more likely to harm than to promote L2 learners’ accuracy development (see Chapters 2 and 4 for a full review of Truscott’s objections against CF). He furthermore stated that the available research base was unable to refute his claim that CF is necessarily ineffective and potentially harmful.

Truscott’s (1996) case against CF gave rise to an ever growing number of studies investigating the effectiveness of error correction. Early CF work mainly focused on the role of CF in the revision process, and demonstrated that correction enables learners to improve the accuracy of a particular piece of writing (e.g. Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001; Sachs & Polio, 2007). Whereas these studies proved that CF is a useful editing tool, they do not validate any conclusions on the role of CF in the process of L2 learning. Investigating the value of CF for L2 acquisition, would necessarily involve “a comparison between two independently written works” (Truscott & Hsu, 2008, p. 293). Revision studies, however, only compare two versions of the same text.

Earlier studies that did opt to investigate CF’s effectiveness in yielding a learning effect (by looking into the effect of CF on newly written texts), failed to provide conclusive evidence on the value of CF for L2 acquisition because of methodological short-comings2 (e.g. Chandler, 2003; Kepner, 1991; Polio, Fleck, & Leder, 1998; Semke, 1984; Sheppard, 1992). A number of recent, tightly controlled investigations, however, has proven that CF does have the potential to foster L2 learning, by showing that CF enables learners to improve the accuracy of new pieces of writing (see Chapters 3 and 4; see also Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007).

Although these recent studies contributed greatly to the error correction debate by tackling the question if CF is an effective means of improving L2 learners’ written accuracy over time, they do not provide insights into how and when learners benefit from error correction, because they have all investigated the effectiveness of written CF by comparing group performances over time, rather than by in-depth comparison of individual learners’

2 See Chapter 3 or, for example, Bitchener (2008) and Guenette (2007) for a review of these design related issues.
sequential accuracy performances. As Storch and Wigglesworth (2010) stated “research that analyzes actual instances of learners engaging with feedback and revising their texts […] is needed to understand how […] learners respond to (different forms of) CF” (p. 304).

Moreover, most of the written CF research – except for the studies investigating the effect of CF on one specific error type (Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010a; Bitchener & Knoch, 2010b; Ellis et al., 2008; Sheen, 2007) – used global accuracy measures to assess CF efficacy (i.e. number of errors per 100 words, number of error-free T-units, etc.). Bruton (2009a) suggested, however, that comparing two texts on such global accuracy measures might not be the best way to investigate if learners benefit from error correction, and argued that more detailed, qualitative analysis of learners’ accuracy performance over time might give a more complete and accurate picture of the accuracy gains brought about by CF. Bruton based this claim on the outcomes of his detailed re-analysis of a portion of Truscott and Hsu’s (2008) data. Comparing the sequential error rates of this study’s CF group (i.e. the experimental group) to those of its control group, led Truscott and Hsu to conclude that the accuracy gained by their CF group during revision did not lead to improved accuracy in a new text. Hence Truscott and Hsu stated that “the successful error reduction during revision [was] not a predictor […] of learning” (p. 299). However, when subjecting a sample of Truscott and Hsu’s data to an in-depth inspection, Bruton did find evidence of a carry-over effect of previous corrections on subsequent writing. He explained that the global error rates used by Truscott and Hsu failed to reveal these observed accuracy gains, because measuring accuracy development by means of global error rate comparison is based on false premises. First of all, it presupposes that a learner will (be able to) use the features that were corrected in his initial text when writing a new one. Secondly, Bruton argued that comparing an initial text and a subsequent piece of writing on global accuracy measures “does not reflect the assumed relationship between the errors in the two texts: it is not that the student writer made some grammatical errors, which were […] corrected in a revision, only to reappear a week later” (p. 139). In reality the sample from Truscott and Hsu’s data showed that a lot of the errors in the subsequent text did not bear any relation to the errors corrected in the first piece of writing.

Another drawback of global accuracy measurement identified by Bruton (2007; 2010) is that, in adopting all-or-nothing criteria (i.e. accurate/inaccurate), it will inevitably fail to quantify partial learning. It is conceivable, however, that a learner will show retention for only part of a particular feedback instance (as will also become clear from our findings).

A final issue into which qualitative accuracy analysis could provide valuable insights,
is the potentially differential effects of various CF forms on distinctive types of errors. It has been argued, that no single form of CF should be expected to be effective in addressing all types of linguistic errors (e.g. Ferris, 1999; Truscott, 1996), because morphological, syntactic, and lexical errors represent gaps within different domains of linguistic knowledge (e.g. Schwartz, 1993). In-depth analyses of learners’ writing would enable exploration of possible interactions between types of errors and CF methodologies.

Two CF methodologies that have received a lot of attention, are direct CF (i.e. both errors and target forms are indicated by the teacher) and indirect CF (i.e. errors are identified without provision of target forms). While some researchers have suggested that learners will benefit more from indirect CF because they have to engage in a more profound form of language processing when they are self-editing their writing (e.g. Ferris, 1995; Lalande, 1982), we (cf. Chapter 4) and others have claimed that the indirect approach might fail because indirect CF provides learners with insufficient information to resolve complex linguistic problems such as syntactic errors (e.g. Bitchener & Knoch, 2010b; Chandler, 2003).

Only two studies that we are aware of, investigated the interaction between error type and the effectiveness of direct and indirect CF methodologies. An explorative study by Ferris (2006) showed that 86 ESL students only realized a significant reduction over the semester in errors concerning verb form, and that the majority of this kind of errors was corrected indirectly. These observations led Ferris to suggest that indirect CF was more beneficial to accuracy development than direct CF. However, this study was not initially designed, and therefore unable to directly compare the two CF methodologies. In fact, Ferris set out to explore the value of indirect CF only, but found that the teachers in her study addressed different types of errors with different forms of CF.

The study reported on in Chapter 4, investigated the effectiveness of direct and indirect comprehensive CF in a tightly controlled classroom-based study (N = 268) incorporating pre-test, treatment, and (delayed) post-test sessions. This study contrasted two experimental treatments (i.e. direct CF and indirect CF) and two control treatments (i.e. self-correction and writing practice). Results not only showed comprehensive CF to be effective in promoting accuracy over time, they also revealed an interaction between error type and CF methodology: whereas direct CF proved to be most effective in remedying grammatical errors (i.e. morphosyntactic errors), learners’ improvement on non-grammatical features (e.g. spelling errors, punctuation errors) was retained the longest when indirect corrections were provided.

See Chapter 4 for a full review of the studies contrasting direct and indirect CF types.
The findings presented in Chapter 4 contribute to the error correction debate (e.g. Truscott, 1996; Ferris, 1999) by refuting Truscott’s (e.g. 2007) claim that grammatical errors are insusceptible to CF, and by showing that direct correction might be most beneficial to learners’ development of grammatical accuracy. A limitation, however, lies in the broadness of the contrasted error categories (i.e. grammatical vs. non-grammatical errors). It might well be that CF types interact differently with separate error types within these broad domains (see also Bitchener, 2008). In-depth analysis of learners’ accuracy performance over time has the potential to reveal such fine-grained interdependencies between the effectiveness of CF forms and error types.

5.3 The present study

5.3.1 Aims
The present multiple case-study adopted detailed, sequential error analyses to explore the effects of written CF on L2 learners’ accuracy development. In doing so, we took the above mentioned reservations towards global accuracy measurement at heart, and acted upon Storch and Wigglesworth’s (2010) call for more qualitative CF studies. With the in-depth approach we aimed at (1) furthering our understanding on how and when L2 learners benefit from written CF, and at (2) providing insights into the (potentially) differential effects of direct and indirect CF on different types of errors. The cases under investigation were taken from the larger quantitative project presented in Chapter 4, in which 268 participants took part.

5.3.2 Setting and participants
From the 268 participants in the larger project, four were selected for closer examination in the present multiple case-study. Participants were pupils in their second year of Dutch pre-vocational secondary education (N = 134) or higher general secondary education (N = 134). They attended schools with multilingual student populations. Over 80% of those schools’ pupils came from non-Dutch language backgrounds; most pupils were born in the Netherlands, but many of them only started learning Dutch in school (i.e. at age four).
All four case-study participants received either direct or indirect CF during the treatment session\(^4\). For each feedback type, we selected one pupil whose error rate decreased after he or she had received CF, and one pupil whose error rate increased between pre-test and post-tests. By selecting pupils who seemingly reacted very differently to the CF they received, we hoped to gain a better understanding of why, how, and when learners (fail to) benefit from correction.

The case-study participants furthermore satisfied the following selection criteria: (i) Dutch was not their L1; (ii) they participated in all four experimental sessions; (iii) they showed an average (within one SD of sample mean) performance on the pre-test measure of written accuracy (i.e. number of linguistic errors per 10 words); (iv) their pre-test writing contained errors within a range of linguistic categories.

### 5.3.3 Treatments and procedure

Participants attended four experimental sessions; a pre-test session (S1), a treatment session (S2), a post-test session (S3), and a delayed post-test session (S4). During the first session (S1) pupils were given 20 minutes to complete the first writing task. They were also presented with a receptive vocabulary test (cf. Appendix E), and a questionnaire concerning their language background (cf. Appendix D). Scores on the vocabulary test provided an indication of pupils’ overall language proficiency (e.g. Zareva, Schwanenflugel, & Nikolova, 2005). During the treatment session (S2), which took place one week after S1, pupils received the corrected versions of their initial text. The first experimental group was provided with direct comprehensive CF, which consisted of identification of all existing linguistic errors and provision of the corresponding target forms. The second experimental group received comprehensive feedback in the form of indirect corrections, consisting of indications of errors and error codes corresponding to the relevant error categories. It was left to the student to derive the target forms. Pupils in both groups were instructed to copy their initial text while revising all errors corrected by the researcher, and were allocated 20 minutes to finish the task. The first post-test (S3) was administered one week after the treatment session (S2), and the delayed post-test (S4) took place one month after S2.

---

\(^4\) Since the quantitative study presented in Chapter 4 already showed that receiving CF was more beneficial to accuracy development than doing without, this case-study’s focus was only on exploring the effects of CF more in-depth. Therefore, the present study does not include pupils who received one of the control treatments (i.e. self-correction or writing practice).
During both post-test sessions, pupils were given 20 minutes to produce a newly written text.

### 5.3.4 Writing tasks

The tasks used throughout the experiment were three writing assignments on the metamorphosis of different insects, that is butterflies (S1, S2), ladybugs (S3), and honey bees (S4). All tasks within the series had a comparable form and set-up; they invited pupils to write an e-mail to a classmate explaining the metamorphosis of the particular insect, based on a series of images depicting the metamorphosis process. (Cf. Appendix A.)

### 5.3.5 Data analysis

As did the studies in Chapters 3 and 4 and other investigations into the effectiveness of written CF (e.g. Chandler, 2003; Truscott & Hsu, 2008), we calculated and compared error ratios (i.e. (number of form-related errors/total number of words) x 10) to be able to measure learners’ accuracy development. Because of the drawbacks associated with this type of global assessment of accuracy improvement (e.g. Bruton, 2007; 2009a; 2010; Storch & Wigglesworth, 2010), however, we also used more detailed measures to establish pupils’ levels of successful CF uptake and retention.

We compared the texts pupils wrote during the pre-test session (S1) with the revised versions produced in the treatment session (S2), along with the provided corrections, to trace for evidence of successful feedback uptake. Utterances in the initial texts on which CF was provided, were identified in the revised versions as either (a) corrected/reformulated correctly, (b) uncorrected, (c) revised/reformulated incorrectly, or (d) deleted. Since successful uptake can be defined as a learner’s correct response to a CF instance, corrections and correct reformulations were taken to show that CF instances were taken up successfully.

Texts written throughout the post-test (S3) and delayed post-test sessions (S4) were analyzed for proof of accuracy development – or learning – brought about by CF. Accuracy development was operationalized as the target-like usage of a corrected feature in a new piece of writing (i.e. texts written during S3 or S4).

Since CF could not be expected to have any effect on features it did not target (Bruton, 2009a), we also considered it important to identify the errors in learners’ revisions.

---

5 See Appendix B for the instructions the different groups received during the treatment session (S2), and Appendix C for a handout on the error coding system pupils in the indirect CF group were provided with.
(S2), post-test texts (S3), and delayed post-test writing (S4) that did not bear any relation to the pre-test (S1) problems they received CF on. We classified such errors as ‘new errors’. It needs to be noted that a relation between two errors does not imply a one-to-one correspondence; when, for example, a learner received feedback on an incorrect agreement marker on a certain verb in his initial text, and he committed the same agreement error on a different verb in a subsequent piece of writing, we did not classify this latter error as a new problem. In this case, we considered the provided CF to be readily transferable to other but similar structures and utterances.

To be able to explore the amenability of different types of errors to CF, the errors in pupils’ writing were classified into four broad linguistic domains, that is morphosyntax, lexicon, orthography, and pragmatics. Within the different linguistic categories errors were subsequently subdivided into 11 separate error types (cf. Table 5.1).

**Table 5.1. Linguistic domains and error types**

<table>
<thead>
<tr>
<th>Linguistic domain</th>
<th>Error type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphosyntax</td>
<td>Word order error</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Omission of a necessary element</td>
<td>Omission of constituent (e.g. subject, object, verb)</td>
</tr>
<tr>
<td></td>
<td>Addition of a non-necessary element</td>
<td>Addition of superfluous constituent</td>
</tr>
<tr>
<td></td>
<td>Determiner error</td>
<td>Omission/superfluous determiner, incorrect determiner</td>
</tr>
<tr>
<td></td>
<td>Referential error</td>
<td>Lack of number/gender agreement pronoun - referent</td>
</tr>
<tr>
<td></td>
<td>Inflectional error</td>
<td>Incorrect verbal, nominal, adjectival inflection</td>
</tr>
<tr>
<td>Lexicon</td>
<td>Word choice error</td>
<td>-</td>
</tr>
<tr>
<td>Orthography</td>
<td>Capitalization error</td>
<td>Omission/superfluous use of upper-case character</td>
</tr>
<tr>
<td></td>
<td>Punctuation error</td>
<td>Omission/incorrect use of punctuation mark</td>
</tr>
<tr>
<td></td>
<td>Spelling error</td>
<td>-</td>
</tr>
<tr>
<td>Pragmatics</td>
<td>Contextual error</td>
<td>Error concerning politeness, pronoun without formerly introduced referent</td>
</tr>
</tbody>
</table>
5.4 Findings

In this section, we present the observations about the accuracy performance of our four case-study participants: Emre, Nathalie, Mehmet, and Dinesh\(^6\). For each of the participants, we first provide some background information, and then consecutively analyze their sequential error rates\(^7\), level of CF uptake, and level of CF retention. We conclude each case-study by summarizing the most important observations.

5.4.1 The case of Emre: indirect CF and decreasing error rates

Background

Emre is a 14 year-old male, who was born in the Netherlands. Both of his parents were born in Turkey, and speak Turkish with Emre. He attended the second year of higher general secondary education at the time of data collection. Emre’s overall language proficiency in Dutch – which was measured by means of a receptive vocabulary test – was below average compared to that of pupils at the same educational level; he answered 64 out of 108 items correctly, while the average score for his level of education was 76.5 (SD = 10.10).

Sequential error rate analysis

Table 5.2 shows the number of errors and their distribution over different error categories in the texts Emre wrote during pre-test (S1), treatment (S2), post-test (S3), and delayed post-test (S4) sessions.

Emre’s initial text (i.e. the text written during S1) contained 23 errors, on which he received indirect CF in the form of correction codes. The overall error rate (i.e. (number of form-related errors/total number of words) x 10) for this text was 2.32.

The sequential error rates of Emre’s texts show that he was able to considerably improve his initial text through revision, and indicate that he still benefited from the CF when writing a new text; at S2 Emre was able to bring back the number of errors in his first text from 23 (S1) to 4, rendering a reduction of 1.51 errors per ten words (i.e. from 2.32 at S1 to

---

\(^6\) The names of the case-study participants reported on in this chapter are fictitious.

\(^7\) It was explained in Chapter 3 that it might not be valid to directly compare pupils’ accuracy performance on different writing tasks, because tasks might differ in difficulty. We believe, however, that the qualitative perspective of the study presented in this chapter allows for such a sequential accuracy analysis; it was adopted to examine learners’ accuracy development in relation to the sequential accuracy performances of the other case-study participants.
0.81 at S2). When comparing the number of errors in Emre’s pre-test text \( (n = 23) \) to his performance on the post-test \( (n = 11) \) and delayed post-test \( (n = 9) \), we see that CF also constituted improved accuracy over time. Emre realized an error rate reduction of 1.36 errors between the pre-test and the two post-tests (from 2.32 at S1 to 0.96 at S3 and S4).

Table 5.2 furthermore shows that 64% of the errors Emre committed at S3 and 44% of the errors he made at S4 were totally new; they did not hold any relation to the feedback provided at S2.

### Table 5.2. Number and types of errors – Emre

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of occurrences S1</th>
<th>No. of occurrences S2</th>
<th>No. of occurrences S3</th>
<th>No. of occurrences S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word order errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omissions of necessary elements</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Additions of non-necessary elements</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Determiner errors</td>
<td>5</td>
<td>1</td>
<td>2 ( (1^N) )</td>
<td>0</td>
</tr>
<tr>
<td>Referential errors</td>
<td>0</td>
<td>0</td>
<td>1(^N)</td>
<td>1( ^N )</td>
</tr>
<tr>
<td>Inflectional errors</td>
<td>3</td>
<td>0</td>
<td>1(^N)</td>
<td>2 ( (1^N) )</td>
</tr>
<tr>
<td>Word choice errors</td>
<td>1</td>
<td>1</td>
<td>1(^N)</td>
<td>1( ^N )</td>
</tr>
<tr>
<td>Capitalization errors</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>0</td>
<td>0</td>
<td>2(^N)</td>
<td>0</td>
</tr>
<tr>
<td>Contextual errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total no. of errors (no. of words)</strong></td>
<td>23 ( (99) )</td>
<td>4 ( (103) )</td>
<td>11 ( (114) )</td>
<td>9 ( (94) )</td>
</tr>
</tbody>
</table>

\(^N\) New error: not present at S1 and no CF provided in relation to this error

### Analysis of feedback uptake

Emre showed a high level of successful feedback uptake; he corrected or correctly reformulated the vast majority of the utterances containing an error (i.e. 19 out of 23, or 83%) during revision (S2). The errors that were left uncorrected are one out of nine capitalization errors, one out of five punctuation errors, one out of five determiner errors, and one word choice error (i.e. inappropriate preposition).

Emre proved to be able to use the indirect CF instances targeting errors in the orthographical domain (i.e. capitalization and punctuation errors), and errors within the domain of morphosyntax (i.e. errors in the use of determiners and nominal inflection). He showed no uptake, however, in relation to the feedback provided on the only lexical error in
his initial text, concerning the use of the incorrect preposition *naar (to) instead of *in (in). The same error reappeared in Emre’s revised text (S2), as can be seen in excerpt (1):

\[(1) \quad \text{S1: veranderen *naar larven}^9 \quad \text{incorrect preposition}\]
\[
\quad \text{change *to larvae}
\]
\[
\quad \text{[change into larvae]}
\]

\[
\text{S2: veranderen *naar larven}^9 \quad \text{uncorrected}
\]
\[
\quad \text{change *to larvae}
\]
\[
\quad \text{[change into larvae]}
\]

This lack of CF uptake might be related to an interaction between the idiosyncratic nature of this lexical error and the indirect feedback Emre received. Ferris (1999; 2002) hypothesized that chances are slim for learners to succeed in self-correcting errors based on indirect CF, when a clear rule behind the solution is lacking. She reasoned that indirect CF provides a learner with too little information to enable him to deduce the target form. Ferris (1999) therefore recommended teachers to give direct CF on such non-rule-based errors, hoping to, “if nothing else, provide input for the acquisition of these idiomatic forms” (p. 6)\textsuperscript{10}.

Based on these suggestions by Ferris (1999; 2002), we propose that, whereas Emre was able to self-edit his rule-based errors concerning capitalization, punctuation, determiners, and inflection, the indirect CF he received failed to provide him with sufficient information to solve his non-rule-governed lexical error in (1).

**Analysis of feedback retention**

Emre’s sequential error rates already showed that he was able to improve the accuracy of his writing over time. However, a closer look at two corrected constructions that reappear in

---

\textsuperscript{8} The asterisk in front of an utterance indicates an erroneous construction.

\textsuperscript{9} In the original text the noun *larve (larva) did not agree in number with the verb *veranderen (to change) (cf. excerpt (3)). To enhance their saliency, however, the excerpts in this chapter only present one error at a time. Errors which were irrelevant for the example at hand were corrected.

\textsuperscript{10} Ferris (1999; 2002) labeled non-idiomatic or idiosyncratic errors as “untreatable” errors. Errors that occur in a patterned and rule-governed way, on the other hand, were categorized as “treatable”. In our opinion this terminology is unfortunate. It seems to propagate the rather fatalistic idea that it is impossible to remedy learners’ non-rule-based errors, irrespective of the type of feedback provided. However, the way Ferris related the nature of errors to their amenability to direct and indirect CF seems to be useful; whereas rule-governed errors may be treatable by indirect CF, idiosyncratic errors appear to be untreatable by indirect correction.
the texts Emre produced during the post-test (S3) and delayed post-test sessions (S4), lead us to interesting observations that stay hidden behind the numbers.

The first construction from Emre’s writing we will examine in more detail, concerns a re-emerging (incomplete or erroneous) determiner-noun combination. In Dutch, nouns in singular contexts must be preceded by an article. Dutch distinguishes between definite and indefinite determiners, and has a two-way gender system (i.e. neuter gender and common gender) which surfaces, for example, in gender agreement between the noun and the definite article; whereas common gender nouns must be combined with *de (the)*, neuter gender nouns are accompanied by the article *het (the)*.¹¹

Excerpt (2) shows that Emre omitted the article which should have preceded the noun *metamorfose (metamorphosis)* at S1. The feedback he received on this error enabled him to correctly revise the utterance at S2. In the text Emre wrote during the first post-test session (S3), he produced the same noun phrase, this time including the article. However, the article *het (the)* and the noun *metamorfose (metamorphosis)* do not agree in gender (the article having neuter, the noun common gender), rendering an error in the determiner category. Finally, during the delayed post-test (S4), Emre did combine *metamorfose (metamorphosis)* with the agreeing determiner *de (the)*.

---

(2)  
S1: Het ging over ø-DET metamorfose-COM SG van een vlinder  
*It was about ø-DET metamorphosis of a butterfly*  
[It was about the metamorphosis of a butterfly]  

S2: Het ging over de metamorfose-COM SG van een vlinder  
*It was about the metamorphosis of a butterfly*  
[It was about the metamorphosis of a butterfly]  

S3: Het gaat over *het metamorfose-COM SG van een lieveheersbeestje*  
*It is about *the metamorphosis of a ladybug*  
[It is about the metamorphosis of a ladybug]  

S4: … vertellen over de metamorfose-COM SG van de wesp  
*… tell about the metamorphosis of the wasp*  
[… tell about the metamorphosis of a wasp]  

---

¹¹ Dutch determiners only distinguish between the two genders in definite, singular contexts. In plural contexts *de* is used in combination with both common and neuter gender nouns. In indefinite contexts all nouns – irrespective of their gender – combine with the article *een (a(n)).*
A comparable pattern of development that is visible from Emre’s writing, concerns the plural formation of the noun larve (larva). Dutch has two options for pluralizing root nouns; the plural form can be created by attaching either the affix –(e)n (/\(\text{n}\)/) or -s (/\(\text{s}\)/). The criteria for determining which affix attaches to which noun, are phonological in nature. One of the relevant features is stress\(^{12}\); whereas a word with a stressed final syllable usually combines with –en, a word ending in an unstressed syllable will prefer –s (Van Wijk, 2002).

Rules governing the plural formation of simplex nouns ending in /\(\text{s}\)/ are less straightforward. While the majority of words in this class correctly combines with both affixes (-n and -s), some exclusively take either -n or -s. As becomes clear from excerpt (3), Emre struggled with the pluralization of such a noun, namely larve (larva). This noun can only be combined with the plural affix -n.

(3)  
\[S1: \ldots \text{"larve-3rd SG zijn-3rd PL ...} \]
\[\ldots \text{larva are ...} \]
\[\text{[... larvae are ...]} \]
\[\text{Ze veranderen-3rd PL in "larve-3rd SG ...} \]
\[\text{They change into "larva ...} \]
\[\text{[They are changing into larvae]} \]

\[S2: \ldots \text{larven-3rd PL zijn-3rd PL ...} \]
\[\ldots \text{larvae are ...} \]
\[\text{[... larvae are ...]} \]
\[\text{Ze veranderen-3rd PL in "larven-3rd PL ...} \]
\[\text{They change into larvae ...} \]
\[\text{[They are changing into larvae]} \]

\[S3: \text{Ze groeien uit tot een-DET SG larve-SG} \]
\[\text{They grow out to a larva} \]
\[\text{[They grow out to be a larva]} \]

\[S4: \text{Ze groeien-3rd PL uit tot "larves-*3rd PL} \]
\[\text{They grow out to * larvae} \]
\[\text{[They grow out to be larvae]} \]

\(^{12}\) Factors such as sonorancy of the final consonant and the number of consonants in the final coda are also involved in Dutch noun pluralization. A comprehensive description of Dutch plural formation is beyond the scope of this thesis, but see, for example, De Haas & Trommelen (1993) for details.
In his initial text (S1), Emre failed to realize pluralization of the noun *larve* (*larva*) twice, resulting in two subject-verb agreement errors; he combined the singular form of the noun *larve* (*larva*) with the plural form of the verbs *zijn* (*to be*) and *veranderen* (*to change*). Just as in example (2), Emre succeeded in producing the target plural form during revision (S2). At S3 he correctly combined the singular noun with the singular form of the verb. In the delayed post-test Emre reintroduced the plural form of the noun *larve* (*larva*), albeit by using an incorrect plural affix –s.

The particular patterns in Emre’s use or failure to use the (correct) determiner and plural marking (i.e. nominal inflection) in excerpts (2) and (3), provide a better understanding of how learners might use indirect CF, and offer insights into the amenability of different types of errors to this kind of feedback. How can it be explained that Emre was able to self-edit the errors in (2) and (3) correctly during revision, failed to use the correct form in one of the post-tests, but produced the target form in the other? To answer this question we again resort to Ferris’ (e.g. 1999, 2002) prediction that rule-based and non-rule-based errors differ in their amenability to indirect CF. In the following paragraphs we suggest that the indirect CF Emre received, only partially remedied the errors presented in excerpts (2) and (3).

Although excerpt (2) shows that Emre committed a determiner error in the same noun phrase before and after receiving CF, the S1 error is different in nature from the error in the post-test text (S3). Whereas Emre used the wrong determiner at S3, he omitted the determiner altogether in his S1 writing. This determiner omission at S1 did not stand on itself. All determiner errors in his initial text were cases of article omission; Emre omitted five out of seven obligatory determiners during the pre-test\(^\text{13}\). Conversely, in his post-test writing, Emre realized all obligatory determiners (i.e. 8 at S3, and 10 at S4). The only remaining determiner errors concerned the use of an incorrect article.

We propose that the indirect CF Emre received, might have raised his awareness of the rule which dictates Dutch singular count nouns to be accompanied by a determiner. It could be assumed, that this awareness then led Emre to add an article during revision, and to not omit any articles in his (delayed) post-test writing. However, although Emre chose the correct article *de* (*the*) when revising, he could not be sure whether his hypothesized choice for this determiner was indeed accurate (see also Chandler, 2003); there is no straightforward rule in Dutch prescribing whether a noun has neuter or non-neuter gender. This uncertainty might have prevented Emre from internalizing the correct form. This

\(^{13}\) The other four determiner omissions concerned different article-noun combinations than the one presented in excerpt (2).
clarifies how it was possible for Emre to incorrectly combine metamorfose (metamorphosis) with het (the) in the post-test, and again correctly with de (the) in the delayed post-test. Whereas the error concerning the rule-governed use of Dutch determiners proved to be susceptible to indirect CF, his error pertaining to the idiosyncratic determiner-noun combination was not. Until Emre is provided with direct input on which article accurately combines with metamorfose (metamorphosis), he has to choose between the two available options (i.e. de and het) every time he produces the particular noun.

Just as the use of determiners, the plural formation of nouns is only partially governed by a clear-cut rule in Dutch, and therefore only partially receptive to indirect CF. In general, simplex nouns with an /s/ ending can take both plural markers -n and -s. Nevertheless, larve (larva) is one of the exceptional nouns which only allows combination with -n. This makes that, for this noun, pluralization is a more idiosyncratic issue, and explains Emre’s sequential performance in excerpt (3). During S1, he failed to use any plural marking on the noun larve (larva). We suggest that – when provided with indirect CF – Emre became aware of the requirement for subjects to agree in number with their matching verbs. Due to this awareness, no omissions of nominal plural affixes remained in the two post-tests. Nonetheless, indirect CF did not succeed in fully remedying Emre’s problems with plural morphology. Parallel to excerpt (2), Emre selected the correct plural marker –n during revision (S2), but the incorrect one (i.e. –s) in the delayed post-test session (S4).

Summary
A comparison of Emre’s sequential error rates revealed evidence of feedback uptake and retention; after receiving indirect CF, Emre produced substantially fewer errors in both his revised text, as well as in his newly written texts. It also showed that about half of the errors in Emre’s post-test writing did not bear any relation to the CF he received. Exploring Emre’s accuracy performance at a more detailed level, however, provided additional insights into the amenability of different error types to indirect CF, and revealed evidence of partial accuracy improvement that was not represented in the overall error rates.

5.4.2 The case of Nathalie: direct CF and decreasing error rates

Background
Nathalie is a 14 year-old female, born in the Netherlands, whose family is from Surinamese origin. Her L1 is Sranan Tongo, a creole language originally spoken by the Creole
population of Suriname. At the time of data collection, Nathalie attended the second year of higher general secondary education. With a score of 91 out of 108 on the vocabulary test, Nathalie’s proficiency in Dutch was above average (i.e. 76.5 items correct, SD = 10.10) in comparison to pupils with the same educational background.

**Sequential error rate analysis**

Table 5.3 presents the number and distribution of errors in Nathalie’s texts. She produced a total of 21 errors in the text she wrote during S1, rendering an overall error rate of 1.71 errors per 10 words. Natalie received direct CF on these errors. The feedback enabled her to reduce the number of errors in her initial text from 21 to 3 during revision (S2). As a result, the overall error rate dropped by 1.47 errors (i.e. from 1.71 at S1 to 0.24 at S2).

A comparison of the number of errors Nathalie committed in her initial text ($n = 21$) to her performance on the post-test (S3) ($n = 7$) and delayed post-test (S4) ($n = 5$), suggests that the direct CF Nathalie received also positively influenced the accuracy of newly written texts; she realized a reduction in error rate of 1.07 errors per 10 words between the pre-test and the first post-test (from 1.71 at S1 to 0.64 at S3), and of 1.21 errors per 10 words between S1 and S4 (from 1.71 at S1 to 0.50 at S4).

Moreover, as illustrated by Table 5.3, 43% of the errors at S3, and 40% of the S4 errors, were new in the sense that they did not bear any relation to the CF Nathalie received.
Table 5.3. Number and types of errors – Nathalie

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of occurrences S1</th>
<th>No. of occurrences S2</th>
<th>No. of occurrences S3</th>
<th>No. of occurrences S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word order errors</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omissions of necessary elements</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1&lt;sup&gt;N&lt;/sup&gt;</td>
</tr>
<tr>
<td>Additions of non-necessary elements</td>
<td>0</td>
<td>0</td>
<td>1&lt;sup&gt;N&lt;/sup&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Determiner errors</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Referential errors</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inflectional errors</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Word choice errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Capitalization errors</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>0</td>
<td>0</td>
<td>2&lt;sup&gt;N&lt;/sup&gt;</td>
<td>1&lt;sup&gt;N&lt;/sup&gt;</td>
</tr>
<tr>
<td>Contextual errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total no. of errors (no. of words)</strong></td>
<td>21 (123)</td>
<td>3 (129)</td>
<td>7 (110)</td>
<td>5 (101)</td>
</tr>
</tbody>
</table>

<sup>N</sup>New error: not present at S1 and no CF provided in relation to this error

Analysis of feedback uptake
As did Emre, Nathalie showed a high level of successful CF uptake; she incorporated 18 out of 21 corrections (i.e. 86%) in the revised version of her initial text. In contrast to Emre, however, Nathalie received direct feedback on the errors she committed. Whereas Emre was presented with the challenging task of correcting his own errors using the error codes in his text, Nathalie only had to copy the corrections provided by the researcher. The fact that she was able to perform this task adequately in 18 out of 21 instances is therefore hardly surprising. One might even question why not all corrections led to uptake, leaving three punctuation errors after revision (two of which were left uncorrected while the other one was revised incorrectly). We suggest that Nathalie might have failed to incorporate two of the provided CF instances due to the relatively low saliency of corrections concerning punctuation, which possibly led her to simply overlook them.

Analysis of feedback retention
Global error rate analysis revealed that Nathalie’s (delayed) post-test writing was more accurate than her initial text. When exploring her accuracy gains across different error categories (cf. Table 5.3), a first observation is that Nathalie showed a lot of improvement in the use of punctuation marks. Her initial text contained 12 punctuation errors: Nathalie used one incorrect punctuation mark, and punctuation marks were missing altogether in 11 out of
28 obligatory occasions. In the text written during the first post-test she committed only three punctuation errors, two of which involved the use of incorrect punctuation marks. At S3 punctuation was failing in only 1 out of 21 obligatory occasions. Finally, during the delayed post-test Nathalie failed to realize punctuation in just two out of 22 obligatory occasions.

Table 5.3 furthermore shows that Nathalie gained accuracy in the grammatical domain; whereas her pre-test text contained seven morphosyntactic errors, only one such error remained in her post-test an delayed post-test writing. The most interesting observation pertains to one of her S1 inflectional errors. A close look at two reappearing constructions in Emre’s writing revealed that non-rule-based errors might be insusceptible to indirect feedback forms. Ferris’ (1999) suggestion that such idiosyncratic errors might benefit more from direct correction seems to be supported by the finding that Nathalie did succeed in acquiring an irregular plural affix based on the direct CF she received.

As explained earlier, Dutch nouns are usually pluralized by adding affixes -s and/or –(e)n. Some particular classes of nouns, however, have to be combined with other markers to realize plural formation. One of these irregular affixes is -eren. The noun blad (leaf) belongs to the class of exceptional nouns that combine with this idiosyncratic affix.

Excerpt (4) shows how Nathalie failed to use accurate plural marking on the noun blad (leaf) in the pre-test (S1). The direct feedback she received enabled her to revise the error correctly (S2). She also succeeded in using the correct form twice in a new text (S3).

(4)  

(4)  

S1: ... vol met *blader-PL ...  
... full of *leaf ...  
[... full of leafs]

S2: ... vol met bladeren-PL ...  
... full of leafs ...  
[... full of leafs]

S3: ... vol met bladeren-PL ...  
... full of leafs ...  
[... full of leafs]

... zich volgegeten met bladeren-PL ...  
... itself stuffed with leafs ...  
[... stuffed itself with leafs]
Summary
Comparing the error rates of Nathalie’s texts supplied evidence of both uptake and retention of direct CF; the revised version of her initial text, as well as Natalie’s post-test writing, contained considerably fewer errors than the piece of writing she produced during the pre-test session. Moreover, more than 40% of the errors Nathalie committed during the post-tests proved to be unrelated to the corrections she was provided with. An in-depth analysis of Natalie’s sequential accuracy performance furthermore highlighted how direct CF succeeded in promoting the accurate use of an idiosyncratic plural affix. Since Emre’s case showed that a similar error was not amenable to indirect CF, we reiterated Ferris’ (1999) suggestion that idiosyncratic errors might best be corrected directly.

5.4.3 The case of Mehmet: indirect CF and increasing error rates

Background
Mehmet is a 14 year-old male, with a Turkish language background; although he was born in the Netherlands, Turkish is the language of communication at home. Mehmet was attending the second year of secondary pre-vocational education when data collection took place. Mehmet answered 55 out of the 108 vocabulary test items correctly, which means that his score, and thus his Dutch language proficiency, are well below average (i.e. 70.8, SD = 10.63) within his educational level.

Sequential error rate analysis
Table 5.4 shows the number and types of errors in Mehmet’s texts. Mehmet committed 23 errors in his initial writing (S1); the overall error rate of this text was 2.25 errors per 10 words.

Contrary to what we saw in Emre’s and Nathalie’s work, Mehmet’s revised text contained more errors ($n = 26$) than his initial version ($n = 23$); there was an increase from 2.25 errors per 10 words at S1 to an error rate of 2.47 at S2. Comparison of Mehmet’s pre-test and post-test performances also reveals that Mehmet committed more errors after receiving indirect CF than before; at S3 the error rate of his text was 2.56, and his delayed post-test writing (S4) held 2.53 errors per 10 words.

When we distinguish between errors that already appeared at S1 and errors that did not, we see that 58% of the errors in the revised text (S2), 52% of the errors in the first post-test (S3), and 50% of the errors in the delayed post-test (S4) are totally new; they hold no
relation to the errors committed at S1, and Mehmet did not receive any feedback on these errors.

Table 5.4. Number and types of errors – Mehmet

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of occurrences S1</th>
<th>No. of occurrences S2</th>
<th>No. of occurrences S3</th>
<th>No. of occurrences S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word order errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omissions of necessary elements</td>
<td>1</td>
<td>1^N</td>
<td>2^N</td>
<td>3^N</td>
</tr>
<tr>
<td>Additions of non-necessary elements</td>
<td>1</td>
<td>1^N</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Determiner errors</td>
<td>0</td>
<td>1^N</td>
<td>1^N</td>
<td>1^N</td>
</tr>
<tr>
<td>Referential errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1^N</td>
</tr>
<tr>
<td>Inflectional errors</td>
<td>4</td>
<td>3 (2^N)</td>
<td>3 (2^N)</td>
<td>4 (1^N)</td>
</tr>
<tr>
<td>Word choice errors</td>
<td>0</td>
<td>0</td>
<td>2^N</td>
<td>0</td>
</tr>
<tr>
<td>Capitalization errors</td>
<td>5</td>
<td>5 (2^N)</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>7</td>
<td>10 (5^N)</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>5</td>
<td>5 (3^N)</td>
<td>4^N</td>
<td>4^N</td>
</tr>
<tr>
<td>Contextual errors</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total no. of errors (no. of words)</strong></td>
<td><strong>23 (102)</strong></td>
<td><strong>26 (105)</strong></td>
<td><strong>23 (90)</strong></td>
<td><strong>20 (79)</strong></td>
</tr>
</tbody>
</table>

^N New error: not present at S1 and no CF provided in relation to this error

Analysis of feedback uptake

When we consider the increase in error rate between the initial version of Mehmet’s text and his revision (i.e. from 2.25 at S1 to 2.47 at S2), the numbers give the impression that Mehmet was unable to interpret and use any of the indirect CF (i.e. error codes) provided to him. However, when we take a careful look at the errors Mehmet committed during S1 and S2, we see that only 11 out of the total 26 errors in his revised text are actually related to errors that were present in the initial version (i.e. the same, unrevised error reappeared, or the initial error was revised incorrectly), and that the other 15 are entirely new. From the 23 utterances containing an error in his pre-test writing, Mehmet was able to correctly revise or reformulate 12 during revision. This leads us to conclude that, while global error rates seemed to show that Mehmet failed to benefit from any of the indirect CF instances at S2, he in fact showed successful uptake for over half of the corrections.

Still, the amount of uptake is relatively low, especially within the orthographical domain; Mehmet left five errors in punctuation, three capitalization errors, and one spelling error uncorrected. One more spelling error was revised incorrectly. As we suggested earlier, the low saliency of (CF instances targeting) punctuation and capitalization errors might
explain the moderate level of uptake within these domains. The lack of uptake concerning the two remaining spelling errors might be related to their persistency, as will become clear from excerpts (6), (7), and (8) in the following paragraph.

**Analysis of feedback retention**

Comparing the error rates of Mehmet's texts would lead to the conclusion that Mehmet did not benefit from the indirect CF he received. To the contrary: his (delayed) post-test writing seemed to be even less accurate than the text Mehmet produced during the pre-test. However, the in-depth analysis of Mehmet's sequential performance leads to two important observations.

To begin with, exploration of the degree to which Mehmet was able to use the indirect CF during revision, brought to light that his uptake of the provided feedback was quite low, especially of CF that targeted orthographical errors; ten out of 17 errors concerning punctuation, capitalization, and spelling were still present in the revised version of Mehmet's text. As it has been suggested that successful uptake might be a predictor of successful retention (e.g. Ferris, 2004), it can be understood why Mehmet's (delayed) post-test writing still contained a lot of errors in orthography.

Secondly, Table 5.4 shows that half of the errors Mehmet committed during the two post-tests hold no relation to the errors he made at S1. The fact that Mehmet's error rate was higher after he received feedback than before his errors were corrected, should therefore not be interpreted as evidence against the usefulness of error correction; CF could not be expected to have any influence on errors it did not target in the first place (Bruton, 2009a).

As explained above, no evidence of learning could be anticipated when either uptake of CF is lacking, or no CF has been provided on a certain erroneous construction. Yet excerpt (5) illustrates that when feedback had been given on a certain issue, and successful uptake in relation to this CF instance had been realized, Mehmet did show indications of feedback retention.

(5)  **S1:** ... kunnen ze *zich vervellen  

       ... can they *REFL shed their skin  

       [... they can shed their skin]

     **superfluous reflexive**  

     **pronoun**

     **S2:** ... kunnen ze vervellen  

       ... can they shed their skin  

       [... they can shed their skin]

     **target form**
Excerpt (5) shows how Mehmet combined the non-reflexive verb *vervellen* (*to shed skin*) with the reflexive pronoun *zich* (*himself*) in his pre-test writing, rendering a syntactic error (i.e. addition of a superfluous element). Subsequently this error was revised correctly, and Mehmet was able to use the corrected utterance adequately at S3.

One more interesting observation we can distillate from Mehmet’s work, relates to his errors concerning inflection and/or (rule-based) spelling. Excerpt (6) shows how Mehmet’s use of the third person singular pronoun *het* (*it*) in combination with the first person singular form (or stem) of the verb *worden* (*to turn into*) resulted in a subject-verb agreement error. Although he corrected this error when revising his text, the same error reappeared four times in his (delayed) post-test writing. In this case, uptake did not translate into retention.

```
S1: Daarna *word-1SG het-3SG een pop
    *turn into it a pupa
    [It then turns into a pupa]

S2: Daarna wordt-3SG het-3SG een pop
    turns into it a pupa
    [It then turns into a pupa]

S3: Daarna *word-1SG een larve-3SG een pop
    *turn into a larva a pupa
    [The larva then turns into a pupa]

S4: ... *word-1SG het-3SG een larve
    *turn into it a larva
    [... turn into a larva]

    ... *word-1SG het-3SG een jonge wesp
    *turn into it a young wasp
    [... turns into a young wasp]

    ... dat je-2SG snel beter *word-1SG
    that you soon well *get
    [... that you get well soon]
```
Why did uptake not result in learning here? The explanation might be found in the nature of the errors in excerpt (6). We propose that instead of labeling them as errors in subject-verb agreement, they could just as well be interpreted as spelling errors. In Dutch the phoneme /t/ can take three different written forms; the graphemes <t>, <d>, and <dt> are all pronounced the same when appearing at the end of a syllable. This makes that the verb forms for first person singular word (turn into) and second/third person singular wordt (turn(s) into) have indistinguishable pronunciations. The incorrect combination of a third or second person singular subject with word (turn into) could therefore be categorized as a spelling error. The errors in excerpt (6) together with several other errors in Mehmet’s texts (cf. excerpts (7) and (8)) suggest that he has a general problem pertaining to the spelling of words that end in the phoneme /t/.

Excerpt (7) shows how Mehmet misspelled two past participles in his initial writing (S1). Both should have ended on <d> rather than on <t>. During revision (S2) Mehmet did not show successful uptake for the CF he received on these errors; he left the first error uncorrected, and incorrectly revised the second one. He then committed a comparable error at S4.

(7)  
S1: … worden *genoemt  
… are *called  
[... are being called]  

… worden *gelegd  
… are *‘aid  
[... are being laid]  

S2: … worden *genoemt  
… are *called  
[... are being called]  

… worden *gelegd  
… are *‘aid  
[... are being laid]  

S4: … je had *gevraagt  
…you had *asked  
[... you asked]
Yet another similar error is presented in excerpt (8). Mehmet committed a spelling error in the diminutive form of the noun *brood* (*bread*); he used a *<t>* instead of the correct grapheme *<d>*.

Excerpts (6), (7), and (8) show that Mehmet’s errors in the use of the different Dutch graphemes representing the phoneme /t/ are very persistent. We propose that such extremely persistent and potentially fossilized (e.g. Selinker, 1972) errors might be to deeply entrenched in a learner’s interlanguage system to be susceptible to CF. This would then explain why feedback uptake was not followed by retention in excerpt (6).

**Summary**

Merely looking at Mehmet’s increasing sequential error rates, would have led to the conclusion that indirect CF harmed rather than improved his written accuracy. A detailed analysis of Mehmet’s performance over time showed, however, that (i) more than half of the errors in Mehmet’s post-test writing did not hold any relation to the errors that were corrected in the text he wrote during the pre-test session, and therefore CF could not be expected to influence those errors; (ii) when CF failed to lead to retention, this was either because no uptake had taken place, or because CF targeted an extremely persistent error; (iii) when CF targeted an error that was less entrenched and the feedback instance was successfully taken up during revision, CF did result in the accurate use of reappearing constructions.

**5.4.4 The case of Dinesh: direct CF and increasing error rates**

**Background**

Dinesh is a 13 year-old male, whose parents are from Surinamese origin. Dinesh was born in the Netherlands, but the language spoken at home is Sarnami Hindustani. At the time of data collection, he attended the second year of secondary pre-vocational education. Answering 70 of the 108 vocabulary test items correctly, Dinesh’s level of proficiency in Dutch is close to average (i.e. 70.8, SD 10.63) within his educational level.
Sequential error rate analysis

The number and distribution of the errors Dinesh committed during the different experimental sessions are displayed in Table 5.5. The text Dinesh wrote during the pre-test session (S1) contained 30 linguistic errors, which translates into an overall error rate of 1.96 errors per 10 words. By revising his work based on the direct corrections he received at S2, Dinesh brought back the number of errors in his text from 30 to 9. Consequently, the overall error rate went down by 1.29 errors per ten words (i.e. from 1.96 at S1 to 0.67 at S2).

Contrary to what we saw in Emre’s and Nathalie’s cases, Dinesh’s error rate development seems to indicate that he was unable to translate the improvement realized during revision into long-term accuracy gains. Instead of showing a reduction in the number of errors between pre-test and post-tests, Dinesh’s error rates were higher at S3 (2.32) and S4 (3.30) than at S1 (1.96). A comparison of the types of errors Dinesh committed during the different experimental stages, reveals that this increase is particularly related to the number of orthographical errors Dinesh made in his post-test writing (i.e. capitalization, punctuation, and spelling errors).

Moreover, 78% of the errors in his revised text (S2), 33% of the errors in the first post-test (S3) and 38% of the errors in the delayed post-test (S4) did not hold any relation to the errors Dinesh committed in his initial text (S1) or received feedback on during S2.

Table 5.5. Number and types of errors – Dinesh

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of occurrences S1</th>
<th>No. of occurrences S2</th>
<th>No. of occurrences S3</th>
<th>No. of occurrences S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word order errors</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Omissions of necessary elements</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Additions of non-necessary elements</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Determiner errors</td>
<td>0</td>
<td>0</td>
<td>2N</td>
<td>1N</td>
</tr>
<tr>
<td>Referential errors</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inflectional errors</td>
<td>3</td>
<td>1N</td>
<td>2N</td>
<td>4 (3N)</td>
</tr>
<tr>
<td>Word choice errors</td>
<td>2</td>
<td>0</td>
<td>1N</td>
<td>3N</td>
</tr>
<tr>
<td>Capitalization errors</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Punctuation errors</td>
<td>10</td>
<td>5 (3N)</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Spelling errors</td>
<td>3</td>
<td>3N</td>
<td>5N</td>
<td>7N</td>
</tr>
<tr>
<td>Contextual errors</td>
<td>0</td>
<td>0</td>
<td>1N</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total no. of errors (no. of words)</strong></td>
<td>30 (153)</td>
<td>9 (135)</td>
<td>33 (142)</td>
<td>37 (112)</td>
</tr>
</tbody>
</table>

N: New error: not present at S1 and no CF provided in relation to this error
Analysis of feedback uptake

As did Nathalie, Dinesh received direct CF. When comparing the number of errors in Dinesh’s initial text \((n = 30)\) to the number of erroneous utterances in the revised version \((n = 9)\) (cf. Table 5.5), it seems as if Dinesh was quite capable of using the feedback he received. Especially when keeping in mind that seven out of the nine errors in his revised text \((S2)\) were not yet present in his pre-test writing \((S1)\), and therefore not targeted by CF. These numbers would lead us to conclude that Dinesh only failed to incorporate 2 out of 30 corrections.

However, a closer look at Dinesh’s revised text brings to light that he actually corrected or correctly reformulated just 17 out of the total 30 errors he committed in his pre-test text \((S1)\). Not only were two punctuation errors left uncorrected, but Dinesh also deleted the utterances containing 11 more errors during revision. These deleted utterances held errors concerning punctuation \((n = 5)\), capitalization \((n = 2)\), omissions \((n = 2)\), word choice \((n = 1)\), and nominal inflection \((n = 1)\). Unlike bare number comparison, a more detailed analysis shows that the level of successful uptake Dinesh displayed was low, especially when taking into account that he received direct CF.

Since gathering process data was not part of the larger research project, we can only speculate on why Dinesh decided to delete some of the utterances he received feedback on instead of correcting them during revision. Earlier studies that did try to tap into learner-internal cognitive processes (by using think-aloud protocols, retrospective interviews, etc.) have shown that feedback uptake is influenced by affective factors such as learners’ motivation, goals, beliefs, and attitudes towards CF (e.g. Bruton, 2010; Goldstein, 2006; Hyland, 1998; Storch & Wigglesworth, 2010; Swain & Lapkin, 2003). We suggest that Dinesh perceived the goal of the revision activity differently from how it was intended by the researcher. Instead of merely copying his initial text and incorporating the corrections, he might have had a broader interpretation of the task as improving his text as he saw fit (see also Storch and Wigglesworth, 2010), deleting and adding complete utterances.

Analysis of feedback retention

Dinesh’s initial text \((S1)\) contained eight sentence level errors (i.e. errors concerning omissions \((n = 6)\), word order \((n = 1)\), and referential expressions \((n = 1)\)). He showed signs of successful uptake for CF instances targeting six out of those eight errors (i.e. he corrected or correctly reformulated the utterances containing these errors during revision \((S2)\)). The utterance holding the two remaining sentence level errors (i.e. two omissions), was deleted from the revised text. It is difficult to trace retention of the corrections targeting
these sentence level errors because the specific corrected constructions do not reappear in Dinesh’s post-test writing. Nevertheless, it is noteworthy that no sentence level errors remain in the texts Dinesh produced during the two post-tests (S3 and S4). Especially the decrease in omission errors (i.e. from six at S1 to zero at S3 and S4) is remarkable; whereas Dinesh omitted a range of different constituents in his initial text (i.e. auxiliary verb, adverb of time, coordinating conjunction, subject, verbal predicate, locative adverb), his post-test writing contained no incomplete sentences.

Another observation pertains to the relation between successful uptake and retention, or learning. Mehmet’s case already showed that a learner will be unlikely to show signs of retention when evidence of CF uptake is lacking. Excerpt (9) below from Dinesh’s text provides further confirmation of this association between the two processes.

(9)  
S1: … komen-3rd PL de *larve-3rd SG … subject-verb agreement error  
…come the *larva …  
[… the larvae come]  
S2: not applicable corrected utterance deleted  
S4: Eerst zijn-3rd PL ze *larve-3rd SG subject-verb agreement error  
First are they * larva  
[First they are larvae]  

Whereas excerpts (3) and (4) showed that errors in noun pluralization such as the one in (9) are (partially) amenable to CF, Dinesh did not benefit from the correction targeting his erroneous plural marking on the noun larve (larva); he committed the same error both in his initial text (S1) and in his delayed post-test writing (S4). The reason behind this lack of retention might be found in the fact that Dinesh did not display any signs of uptake in relation to the feedback instance; instead of correcting or reformulating the S1 utterance in (9), Dinesh chose to delete it from his revised text (S2). We hypothesize that, although the direct CF Dinesh received had the potential to treat the pluralization error, it failed to do so because of the lack of uptake at S2, successful uptake having been proposed to be a “necessary step in developing longer term linguistic competence” (Ferris, 2004, p. 56).

In the same line of reasoning, we propose an explanation for Dinesh’s inability to improve his orthographical accuracy between the pre- and post-tests. Uptake of CF in the orthographical domain was relatively low; during revision (S2) Dinesh only succeeded in correcting 5 out of 10 erroneous uses and omissions of punctuation. This lack of successful
CF uptake might account for the fact that Dinesh showed no accuracy gains over time in the orthographical domain.

**Summary**

Analysis of Dinesh’s sequential error rates seemed to show that he was relatively successful in incorporating the direct corrections he received during revision, but that this accuracy improvement was not retained in new pieces of writing. In-depth text analysis, however, revealed a different image of Dinesh’s accuracy development. We observed that (i) Dinesh deleted a considerable number of corrected utterances from his text during revision, and that, (ii) consequently, his level of successful CF uptake was relatively low; (iii) this limited degree of uptake translated into low levels of retention; and (iv) when CF was successfully taken up – as in the case of corrections targeting sentence level errors – Dinesh did show long-term accuracy improvement in relation to the provided CF.

### 5.5 Conclusions, discussion, and implications

It has been suggested that global measures of accuracy improvement (e.g. error rate comparison) might fail to give an adequate and complete picture of the effects of CF on L2 learners’ writing (Bruton, 2009a). Hence, we subjected the sequential writing performances of four L2 learners of Dutch to an in-depth accuracy analysis. In doing so the present multiple case-study aimed at (1) advancing our understanding on how and when written CF (fails to) affect(s) L2 learners’ accuracy development, and at (2) expanding insights into the amenability of different error types to direct and indirect CF.

Our findings proved that in-depth error analyses could indeed be considered a valuable supplement to the more common quantitative method of assessing accuracy development. We found that the detailed approach provided both different and additional information as compared to global accuracy measurement.

To begin with, the following three observations lead us to conclude that the effectiveness of CF is underestimated by overall error rate comparison: (i) For all four participants their progress, as measured by global scores only, was partially masked by the new errors they committed in their subsequent writing; around 50% of the errors in pupils’ (delayed) post-test texts was not related in any way to errors in their pre-test writing and the CF they received on those S1 problems. As Bruton (2009a) rightly argued, CF could not be expected to promote the accurate use of features it did not target. (ii) Moreover, Emre’s
writing showed that, in certain cases (cf. excerpt (2) and (3)), CF might lead to partial acquisition of a corrected feature. This type of improvement, however, will not be represented in global accuracy scores, since it is only the nature and not the number of errors that changes (Bruton, 2007; 2010). (iii) Finally, whereas in Mehmet’s and Dinesh’s cases overall error rate comparison seemed to show that CF harmed their accuracy development rather than promoting it, in-depth analyses revealed that – under the right circumstances – correction did trigger long-term accuracy improvement.

The detailed error analyses we performed, did not only yield different results from global accuracy measurement. As we aimed for, it also provided us with additional insights concerning the effectiveness of error correction. Regarding this study’s first objective – exploring when and how CF might be beneficial to L2 learners’ accuracy development – we were able to make three interesting observations.

First of all, our findings suggested that the level of successful CF uptake might be predictive of a pupil’s success in acquiring a target form; analyzing the texts written by Mehmet and Dinesh, we observed a lack of feedback retention in those cases where the pupils failed to uptake the provided corrections during revision. This finding is in line with the suggestion in the literature that successful uptake might play a facilitative role in acquisition (e.g. Chaudron, 1977; Lightbown, 1998; Loewen, 2004; Sheen, 2004; Swain, 1985), or could even be a prerequisite for learning (e.g. Ferris, 2004). The idea behind this theorized relation between uptake and retention is that noticing is a necessary condition for acquisition (Schmidt, 1995), and that uptake might be interpreted as (one of the) manifestation(s) of noticing.

A second observation pertaining to the relation between uptake and learning is that successful CF uptake is no guaranty for long-term acquisition of the targeted feature. As excerpt (7) from Mehmet’s work showed, CF targeting an exceptionally deep-rooted and potentially fossilized error might not result in retention, even though the correction was taken up during revision.

Thirdly, in-depth analyses of learners’ errors and corrections provided us with some hints on why pupils sometimes failed to take up the CF they were provided with. The majority of the corrections which did not lead to successful uptake targeted punctuation and capitalization errors. We argued that our learners might have failed to incorporate these feedback instances because of their low saliency; not only their small (physical) size, but also the fact that errors of these types do not greatly affect the comprehensibility of a text, make them relatively easy for learners to overlook. Moreover, we suggested that the level of successful CF uptake might be influenced by affective factors, such as the way in which
learners perceive a revision task (e.g. Storch & Wigglesworth, 2010). We found that Dinesh added and deleted whole utterances (including ones he received CF on), and hypothesized that his interpretation of the revision’s goal might have been broader than copying his initial text and incorporating the corrections as the instruction prescribed. By deleting corrected sentences, his (potential) level of CF uptake decreased.

The second aim of our study was to find out if all errors are equally amenable to direct and indirect CF. Firstly, our findings lead us to believe that – apart from errors which are too entrenched in a learner’s interlanguage grammar to be susceptible to CF (cf. excerpt (6), (7), and (8)) – all types of errors can benefit from correction; our participants showed to be able to gain accuracy on a broad range of linguistic features (e.g. determiner usage, agreement, inflection, punctuation, capitalization, and sentence structure), both on a local level (i.e. within one constituent) and a more global, sentential level. What is important to note, is that the improvement brought about by CF is not item-based; the case of Emre, for example, showed that the CF he received on the article omissions in his initial text led him to realize all obligatory determiners in his post-test writing, not just the ones that were corrected before.

While all investigated error types proved to be correctable, we did find that different types of errors vary in their responsiveness to indirect CF. We observed, for example, that the indirect CF Emre received led him to insert a determiner in front of each noun in his post-test writing, as Dutch grammar prescribes for definite contexts (cf. excerpt 2). However, the indirect corrections left him with insufficient clues to self-edit his idiosyncratic lexical error in excerpt (1). Excerpt (4) from Nathalie’s text, on the other hand, illustrated how direct CF did lead to accuracy development of an idiosyncratic feature. Based on these observations – and following Ferris (1999) – we proposed that, whereas rule-based errors seem to be suitable candidates for self-correction based on indirect CF, direct CF might be more beneficial to learners’ errors that are not rule-governed. This proposition is in line with DeKeyser’s (2003) suggestion that arbitrary form-function mappings will require more explicit learning processes. Although indirect CF is explicit in its corrective nature, it is less explicit than direct correction because it does not provide a learner with the target form. In those cases where a learner cannot depend on a clear rule when trying to deduce a target structure (e.g. when revising a lexical error), indirect CF might not be explicit enough to be beneficial. Therefore, maximally explicit feedback methodologies – such as direct CF –

---

14 Evidence of learning might be harder to find with respect to CF instances that are necessarily targeted at a specific item only, such as corrections of lexical and (most) spelling errors; chances are small that the corrected items can be used in a new text.
could be expected to be most effective in developing the accurate use of idiosyncratic features.

Even though the suggestions in this chapter are based on the individual performances of only four L2 learners, and one should be cautious generalizing the findings, we believe the present research added to our understanding on how and when written CF is beneficial to accuracy development. The detailed analyses adopted in the study provided valuable insights into the relation between CF uptake and acquisition, and the amenability of different types of errors to direct and indirect error correction, which would have stayed hidden behind the numbers in a quantitative approach.