Differential effects of corrective feedback on learner errors

Ioanna Karagianni
61st Primary School of Thessaloniki
joannakar@gmail.com

Abstract
This paper reports on a follow-up study of primary school English language learners’ oral production of comparative, superlative forms after a short period of treatment which involved corrective feedback (CF) in the form of prompts, in L1, L2 or no CF at all. What is particular to this study is the focus on learner errors and their patterns as a measure of L2 learning, which, immediately after treatment in short-term, varied according to CF and learners’ level and gave CF lows an advantage, whereas in the long run the gains behind the errors pertained to learner level only.

Keywords: corrective feedback, prompts, teacher L1/L2 use, errors, acquisition

1 Introduction

The first aim of the present paper is to investigate the effect of teacher oral corrective feedback (CF) on learners’ specific grammatical development in English as a foreign language (L2) from the perspective of learner errors, whereas the second aim is to explore the role of the teachers’ use of learners’ first (L1) or L2 language in this. For this reason, among the teacher moves that are identified as responsive to learners’ erroneous utterances or indicative of corrective intention (e.g. by Lyster & Ranta 1997) only those that lend themselves to L1 will be addressed.

Specifically, the focus will be on prompts, a superordinate category of CF, which, to their implicit end of corrective function and use of metalanguage, involve repetition of learners’ errors, clarification requests, whereas explicitly they take the form of metalinguistic cues or comments and require self-repair (Li 2010; Lyster
Effects of feedback on learner errors

2004; Ranta & Lyster 2007; Sheen & Ellis 2011). Prompts are also usually juxtaposed with recasts, which linguistically non-overtly reformulate learners’ utterances, but their acquisitional value, attested from different strands of CF, was referred to in relation to error-free rather than erroneous production.

2 The role of learner errors and teacher CF in second language acquisition

Insofar as teacher CF responds to learner errors, whose count is usually intended as a measure of CF success, it is worth reviewing the criteria and the theoretical perspectives of error. According to Corder (1981) and Long (1991), errors are deviant language learning forms that are distinguished from mistakes on the grounds that they are systematic, persistent, frequently occurring, reflecting gaps in the learners’ underlying L2 system. Regarding their source, this varies in second language acquisition (SLA), as in building their L2 system, learners may make transfers from their L1 or they may utilize overgeneralization, simplification or omission of L2 rules, reflecting thus respectively interlingual errors or developmental processes common in L1-L2 acquisition (Ellis 1990, 1991; Lightbown & Spada 1999). In light of these non-judgmental perspectives of learner error, error analysis (EA) also exists with the aim of tracing, describing and catering to learner errors and processes. As Corder (1981) succinctly puts it, learner errors carry an added pedagogical and theoretical value because they inform of the progress and processes of SLA.

Teacher CF is also a stepping stone to SLA, in that it accomplishes a number of relevant functions and objectives, beyond and above that of mere correction, deriving support from key concepts in SLA. First, drawing on skill acquisition theory, which places emphasis on practice and views language learning as a gradual movement from more to less controlled and automatic processing (Anderson 1983, 2005; Johnson 1996), and considering the tenets of Swain’s (1985, 1993) Output Hypothesis many scholars (e.g. DeKeyser 2007; Li 2010; Lyster 2004; Mackey 2007; McDonough 2005) argue that the value of CF lies in its capacity to promote, along with production practice, hypothesis testing and revision, by helping learners adjust, rehearse their responses and consolidate error-free production. Besides, cognitively speaking, with reference to Schmidt’s (1990, 2001) Noticing Hypothesis, learners gain greater awareness and understanding of both the nature of their error and the target form.
The evidence that lend support to this theoretical argumentation in favour of teacher CF, particularly prompts, come from different strands of CF research, observational, experimental, perception/reflection, but those assumed relevant to the present come from the studies of Ammar & Spada (2006), Ellis (2007), Ellis et al (2006), Li (2013), Loewen & Nabei (2007), Lyster (2004) because these are the classroom-based, hence of greater external validity and all involve a comparative examination of prompts in relation to no CF or prompts. Generally, their common finding is the superiority of prompts not only over no CF but also over recasts in that they are usually associated with higher and more sustainable post-test scores, suggesting that prompts are better at involving learners into the processes of output modification and noticing. Equally important, they have pinpointed key factors in the success of CF, such as learners’ proficiency level (Ammar & Spada 2006), type of form (Ellis 2007), duration of treatment (Loewen & Nabei 2007). Yet, all have focused on correct responses and took place in higher education ESL settings.

3 The use of L1 in instructed language settings

The study of teacher L1 use within the framework of CF literature is relevant because, by delivering CF, the former realizes, serves and assumes some of the latter’s functions. For instance, in Li’s (2013) CF study, it was decided that metalinguistic explanations would be in L1 to ensure comprehension, whereas in observational L1 use studies (Nakatsukasa & Loewen 2015) similar patterns were obtained. Still, the language of CF has not been isolated as a variable probably because the focus has been on the efficacy of prompts versus recasts, namely on the explicit-implicit dimension and the output-pushing, input-providing aspect of CF (Ellis 2006). Besides their scope, most CF studies took place in ESL settings where learners’ L1 may not overlap. Thus, given the appropriate context, the Foreign Language (FL) and a common L1, the L1/L2 in CF can be treated as an additional variable corresponding to calls for more research into criterial features within the same type of CF (Ammar & Spada 2006; Lyster 2004; Lyster & Ranta 2013).

As for the arguments posited in favour and against the role of teacher L1 use, the former, drawing on socio-cognitive theories, embrace teacher L1 use as a tool that takes into account and corresponds to learners’ natural tendency to use L1 to manage
their L2 learning process (Macaro 2005; Stern 1992; Storch & Aldosari 2010; Swain & Lapkin 2000). In contrast, the opposing ones, by prioritizing the quantity, quality and modification of L2 input delivered in the context of interaction (Krashen 1985; Long 1983), stress that L1 use may undermine the place of L2 in class by assuming some of its functions. Yet, in the absence of experimental research into teacher L1 use, any of these assumptions remain open to question.

Last but not least, irrespective of any theory model, in actual practice teacher L1 use is often utilized as a time-saving device or it is evoked by the orientation of the lesson being higher in form-focused rather than in meaning/communication-based ones (Ellis & Shintani 2013). Also, another determining factor in teacher L1 use seems to be the learners’ proficiency level, as when this is limited, in beginner, low-intermediate or weaker students, this tends to be higher to ensure understanding (de la Campa & Nassaji 2009; Polio & Duff 1994). That way teacher CF and L1 use share not only common functions and conceptual frameworks but also common factors.

### 4 English comparative and superlative

In order to obtain measurable and comparable results as to L2 development, experimental CF research has focused on the effect of CF on the acquisition of particular L2 structures. In one such study of the effect of recasts and metalinguistic feedback, Ellis (2007) targeted the learning of the comparative suffix of monosyllabic adjectives ‘-er’ in relation to regular past tense ‘-ed’, arguing that the former is more difficult than the latter in that it requires attention, to morphological, syntactical features, it relates to other constituents of a sentence and it is less frequently occurring. On top of that, he identified a number of common errors among ESL learners in the use of comparative structure, such as the omission of its free (than) or bound morpheme (-er), the double marking of a form (-erer), the misformation.

With regard to the superlative structure of monosyllabic adjectives, in the absence of clear evidence from SLA research, findings from first language acquisition are utilized. Specifically, drawing on linguistic evidence, Layton & Stick (1979) argue that comparative tends to be acquired first because it occurs more frequently, tends to be produced more frequently by learners and poses less conceptual demands on learners in that it requires two dimensions be attended in a comparison. So, in a sense,
their report of overuse or of overgeneralization of the comparative suffix ‘-er’ prepares the ground for the possibility of similar patterns in SLA.

5 The present study

Considering the previous findings in the use of comparative and superlative structures as well as the research gaps identified in the combined effect of CF and teacher L1, the following research questions (RQ) were formulated:

RQ 1: Does CF, in the form of prompts, aid the L2 learning of the comparative and superlative structure? If so, is there an advantage of CF-L1 over CF-L2?

RQ 2: What types of errors do Greek learners of English commit in the use of comparative and superlative structure of regular monosyllabic adjectives? Is there a difference in the patterns according to their level, CF condition and time of testing?

RQ 3: Does L1-L2 in CF affect the type of learner errors differentially?

As regards RQ1, from the CF research, it was expected that CF groups would benefit more, whereas from the arguments posited in favor teacher L1 use, it was assumed that students of lower proficiency would benefit more from CF-L1. With respect to RQ2, errors of omission, wrong addition were anticipated at the time of pre-test mostly, whereas errors of overgeneralization, substitution or double marking of the target suffixes were anticipated at the time of post-tests. In the absence of relevant evidence, no predictions were made for RQ3.

5.1 Participants and conditions

The subjects of the present study, 30 in total, came from the same context, a model experimental primary school of Thessaloniki, Greece, which as to its orientation of English teaching was meaning-based. These were fourth graders, aged 9 to 10, and of pre-intermediate general English proficiency on the basis of the school placement test. Also, all shared Greek as L1. Another selection criterion was their no or little prior as well as their no meantime exposure to the target structures, except for that they would receive during the instructional intervention which was arranged for the purposes of this study only to take place at three separate classes, each forming one condition.
Specifically, depending on their class/condition during the intervention, one sample/group of students would receive CF in L1 (N=10), the other would receive the same type of CF in L2 (N=10), while the control group would receive NO-CF (N=10). Based on their school placement test scores, these subjects were also grouped into higher and lower proficiency, so that within each CF group two sub-groups of higher and lower proficiency existed totaling in six sub-groups of 5 students each: CF-L1 (5 highs, 5 lows), CF-L2 (5 highs, 5 lows), NO-CF (5 highs, 5 lows). Finally, in order to maintain consistency in the amount, type and language of CF, the researcher, namely the author of this paper, and not the school teacher was the instructor of all groups.

5.2 Operationalisations and materials
In order to isolate the effect of CF, L1/L2 and ensure consistency in the amount and type of practice and production across all conditions/classes, all groups, after a rule of thumb in the beginning, necessitated by the novelty of the structures, received the same type of instruction, materials, and followed the same testing procedures, differing only as to the provision of CF. It was crucial then for the instructor, during the instructional intervention with each group, to be consistent in terms of CF, which meant giving no CF to the errors the NO-CF group committed, while offering only prompts to those of the CF-L1, CF-L2 groups, in L1 or L2 accordingly, through repetition and metalinguistic cues. The following extracted episode exemplifies prompts, with the part that could be delivered in L1 or L2 underlined:

(1)  S: The lion is strongest animal in the jungle.
    T: Only strongest? You need something before strongest for the superlative.

Overall, the organization of the teaching material followed the Presentation-Practice-Production model and was spread over a 70-minute instruction for each CF condition, which in Norris & Ortega’s (2000) account is considered short. At the stage of presentation, the rules of thumb for the target structures were given, whereas the stages of practice and controlled production included two meaningful oral drills, three gap-filling activities and one picture description task, aiming at the comparison of items of the same category with the aid of pictures (e.g. animals, means of transport). Given the research objectives and the shortness of the instruction, only oral
production was tapped and the instructor’s CF, in L1 or L2, targeted only oral learner errors in comparative-superlative, which in the NO-CF condition were ignored.

In the same vein, for the elicitation of the oral test data that were the basis of the present study picture description tasks resembling those of the instruction, on the basis of 5 pairs of pictures of items, targeting comparative, and one card of 5 items of the same category, targeting superlative, were developed in order to obtain comparable controlled production of either target structure from all groups. At every testing time, these tasks addressed different real adjectives and were accompanied by 8 cards that required the comparison of weird items with non-existent, NONCE forms, taken from Agathopoulou (2009), Agathopoulou & Papadopoulou (2009a, 2009b), in order to address rule-based learning. In other words, across all sessions, the test format was the same, but it differed as to the adjectives tested from time to time. Finally, it took 15 minutes to complete and to each group was administered 2 weeks prior to instruction for pre-test, 1 day after it for immediate post-testing and 1 month after it for delayed post-testing, with the aim of making any possible or sustainable gains traceable.

5.3 Coding of errors and data analysis
Given the format of the oral test, the learners’ correct responses were scored out of 18 (out of 9 for comparative and 9 for superlative), totaling thus in a maximum of 540 correct responses, whereas the number of learners’ erroneous responses was calculated in frequencies per CF group and level. To obtain the further distribution of these learner errors across different categories, errors referred to as omission in Ellis (2007) were coded as ‘only -er/-est’ if ‘than’, ‘the’ were not used (e.g. the horse taller sheep) or ‘base’ if both the target suffixes and the bound morphemes were absent (e.g. the horse tall of the sheep). Those of wrong addition and formation were coded as ‘other suffixes’ (e.g. this man is bripps of all), whereas instances of overgeneralization or of double marking of either form were coded as ‘alter -er/-est’ or ‘both -er/-est’ (e.g. this thing is the plimmerst of all).

All analyses were quantitative and the statistical test that best suited this grid of learners’ level, CF condition and testing time, as between and within-subjects factors respectively, was the mixed design factorial ANOVA, within the frame of General Linear Models and with the criterion of Least Significance Difference. The significance level was set at p<.05 and all analyses were conducted in SPSS 21.
6 Results

An overview of learners’ correct responses in Table 1, labeled as target, indicate that generally across all CF conditions the learners’ oral performance was almost the same having elicited small differences. The only exception to this pattern seems to be NO-CF highs and NO-CF lows in immediate post-test, who fell behind CF-L2 highs by 24% and CF-L1 lows by 21% respectively. Even so, no statistically significant group difference was obtained, hence no ostensible superior effect of CF, given the lack of the Test x Level x Group interaction [F(4, 48)=.911, p=.979, partial-\(\eta^2=.01\)].

<table>
<thead>
<tr>
<th>group/level</th>
<th>pre-test</th>
<th>immediate post-test</th>
<th>delayed post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>target</td>
<td>non target</td>
<td>target</td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF-L1</td>
<td>(19/90)</td>
<td>(71/90)</td>
<td>(68/90)</td>
</tr>
<tr>
<td>CF-L2</td>
<td>(18/90)</td>
<td>(72/90)</td>
<td>(76/90)</td>
</tr>
<tr>
<td>NO-CF</td>
<td>(17/90)</td>
<td>(73/90)</td>
<td>(54/90)</td>
</tr>
<tr>
<td>total high</td>
<td>(54/270)</td>
<td>(216/270)</td>
<td>(198/270)</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF-L1</td>
<td>(6/90)</td>
<td>(84/90)</td>
<td>(49/90)</td>
</tr>
<tr>
<td>CF-L2</td>
<td>(5/90)</td>
<td>(85/90)</td>
<td>(44/90)</td>
</tr>
<tr>
<td>NO-CF</td>
<td>(6/90)</td>
<td>(84/90)</td>
<td>(30/90)</td>
</tr>
<tr>
<td>total low</td>
<td>(17/270)</td>
<td>(253/270)</td>
<td>(123/270)</td>
</tr>
</tbody>
</table>

*Table 1. Oral Production: Frequency in % of learners’ correct and erroneous responses per CF group, level and in total over time*

The uniformity of learners’ oral correct performance, which was initially extracted, necessitates the analysis and the categorization of learner errors, as these can be more informative of the learners’ progress. It is noted that among all differences only the significant group differences, which directly correspond to the research questions, are
reported. To begin with the types and the distribution of the learner errors before treatment (Table 2), as expected given the novelty of the structures, the use of base forms was the most frequent error for all, followed by the use of other suffixes especially among lows. The significant Error x Level interaction [F(4, 96)=3.373, p=.013, partial-η²=.123] confirmed the error variation per level, hence that the lows’ errors significantly outnumbered the highs’ errors in this area.

<table>
<thead>
<tr>
<th>Level</th>
<th>Group</th>
<th>only -er/-est</th>
<th>alternation</th>
<th>base forms</th>
<th>other suffixes</th>
<th>both -er/-est</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>CF-L1</td>
<td>4 (9/216)</td>
<td>9 (19/216)</td>
<td>15 (32/216)</td>
<td>4 (9/216)</td>
<td>1 (2/216)</td>
</tr>
<tr>
<td></td>
<td>CF-L2</td>
<td>5 (10/216)</td>
<td>10 (21/216)</td>
<td>15 (33/216)</td>
<td>2 (5/216)</td>
<td>1 (3/216)</td>
</tr>
<tr>
<td></td>
<td>NO-CF</td>
<td>4 (8/216)</td>
<td>5 (11/216)</td>
<td>19 (42/216)</td>
<td>3 (7/216)</td>
<td>2 (5/216)</td>
</tr>
<tr>
<td>total high</td>
<td></td>
<td>13 (27/216)</td>
<td>24 (51/216)</td>
<td>49 (107/216)</td>
<td>10 (21/216)</td>
<td>5 (10/216)</td>
</tr>
<tr>
<td>Low</td>
<td>CF-L1</td>
<td>3 (9/253)</td>
<td>3 (8/253)</td>
<td>12 (30/253)</td>
<td>12 (31/253)</td>
<td>2 (6/253)</td>
</tr>
<tr>
<td></td>
<td>CF-L2</td>
<td>2 (4/253)</td>
<td>2 (5/253)</td>
<td>21 (54/253)</td>
<td>7 (18/253)</td>
<td>2 (4/253)</td>
</tr>
<tr>
<td></td>
<td>NO-CF</td>
<td>2 (5/253)</td>
<td>-</td>
<td>21 (53/253)</td>
<td>9 (24/253)</td>
<td>1 (2/253)</td>
</tr>
<tr>
<td>total low</td>
<td></td>
<td>7 (18/253)</td>
<td>5 (13/253)</td>
<td>54 (137/253)</td>
<td>28 (73/253)</td>
<td>5 (12/253)</td>
</tr>
<tr>
<td>total error type</td>
<td></td>
<td>10 (45/469)</td>
<td>14 (64/469)</td>
<td>52 (244/469)</td>
<td>20 (94/469)</td>
<td>5 (22/469)</td>
</tr>
</tbody>
</table>

Table 2. Distribution of the learner errors in % per CF group, level and in total at the time of pre-test

If the increase of learners’ oral correct responses over time, depicted in Table 1, indicates L2 progress after the instruction, the change in the distribution of the learner errors in Table 3 confirms it and exemplifies it, as the prevalence of the application of the target suffixes only and their alternation supersedes the use of other suffixes. Besides, a combined superior effect of CF per level appeared given the significant Error x Level x Group interaction [F(8, 96)=2.107, p=.042, partial-η²=.15]. The post hoc analyses clarified that NO-CF highs’ errors of the target suffix only (36% of all highs’ errors) significantly exceeded those of CF-L2 highs (p=.033), while NO-CF lows utilized base or other forms significantly more frequently than CF-L1/L2 lows (p=.012, p=.055) accounting for 20% (12% and 8% respectively) of all lows’ errors.
The maintenance of post-treatment gains is displayed in Table 1 overall and in Table 4 in detail. From Table 1, it seems by the time of delayed post-test some treatment gains faded out, especially among lows, as their erroneous productions outnumbered the correct ones. As Table 4 illustrates, in sum the loss of gains pertained to the reappearance of base forms at a rate (20%), comparable to the alternation of the target suffixes (28%), as well as to the reemergence of other suffixes in lows only. In contrast, the pattern of highs’ errors was more stable and the significant Error x Level interaction [F(4, 96)=4.334, p=.003, partial-η²=.15] and post-hoc analyses confirmed the presence of the effect of level as in pre-test, with the sum of lows’ base and other forms reaching higher than that of highs (p=.016, p=.001).
Table 4. Distribution of the learner errors in % per CF group, level and in total at the time of delayed post-test

To summarize the main results, the most frequent errors the learners in general produce after treatment is the use of the target suffixes only, namely without their free morphemes, or the alternate use of them. Regarding the use of base forms or other suffixes, this is affected by time of testing and learners’ level primarily and to lesser extent by CF, as this pertains to the low level more before and long after the treatment and only to NO-CF lows immediately after the treatment. All in all, in some respect, certain treatment gains that were initially obscured by frequency counts or surface comparisons of learner correct or overall erroneous production are revealed.

7 Discussion and conclusions

Reviewing the main findings, in response to RQ1, as to error-free oral production, it can be said that CF in the form of prompts aids the acquisition of comparative and superlative, as better performance is obtained after treatment, but its effect is not differential enough from that of the instruction without CF. The choice of L1/L2 in CF does not seem to play a differential role either given the non-significant group
differences. This seeming neutralization of CF, which was also reported in Lyster’s (2004) study, can be attributed to the rule explanation and the intensive practice the learners of all conditions were given for the linguistic targets, which compensated for the lack of CF. This entails practice by itself turns to a learning process by offering opportunities for greater control over new and existing knowledge via hypothesis testing, output production and modification (DeKeyser 2007; Mackey 2007).

Regarding RQ2, except for the use of other suffixes, most error types (i.e. alternation of ‘-er/-est’, or both) coincided with those of overgeneralization, double marking Ellis (2007), Layton & Stick (1979) reported in L2/L1 acquisition respectively for these structures. This suggests some errors, especially in recently taught forms, are universal across not only L2 learners but also L1-L2, being of developmental nature and indicative of patterns of acquisition mostly, which by and large instruction, including CF, cannot alter or overcome (Ellis 1990, 1991, 1994). Instead, the quality, speed or durability of learning is more amenable to instruction (Spada 1997), as at least at the low level immediately after the treatment the CF group errors of the use of only ‘-er/-est’ or of the alternation of ‘-er/-est’ resembled target-like forms and outnumbered the respective ones of NO-CF, whereas among all highs mainly these prevailed. This also means that, when treatment effects fade out or prior to intervention, the learners’ level of proficiency and readiness is more influential than CF, as the higher it is the more resources learners have to cope with different kinds of stimuli (Ammar & Spada 2006; Li 2013). Thus, returning back to the second part of RQ2, time and learners’ level relate to the patterns of errors obtained, as in the end the highs made greater use of the target suffixes only than the lows.

Finally, with regard to the effect of the choice of the language of CF on the type of learner errors, addressed in RQ3, this was reported to be non-differential, as at no testing time, at any level, no significant group difference between CF-L1 and CF-L2 groups was reported and this applied to learners’ error-free production, too. That is to say, the aforementioned patterns of learner errors did not differ according to the language of CF, as in them the effect of level, testing time and, temporarily, of CF prevailed. This means that in the context of intensive grammar practice L1/L2 may not make a difference, whereas as to the short-lived superior effect of CF among lows it should be noted that, setting aside all factors, the short duration of the instruction itself might be insufficient to trigger big changes (Loewen & Nabei 2007).
As a concluding remark, it is worth mentioning that, despite its limitations as to the small sample size or the shortness of intervention, the present study, in the fields of CF and teachers’ L1 use, contributed to shifting attention from error-free production to the patterns of learner errors as a measure of success of any kind of explicit treatment. Yet, it remains open to question and appealing for future research what patterns would be obtained in relation to familiar, partially established forms or under implicit conditions of teaching to young, as in this study, or older students.

References


