# Evolution of Self-Repair Behaviour in Narration Among Adult Learners of French as a Second Language

Daphnée Simard
*Université du Québec à Montréal*

Leif French
*Sam Houston State University*

Michael Zuniga
*Université du Québec à Montréal*

**Abstract**

*Self-repairs*, or revisions of speech that speakers themselves initiate and complete (Salonen & Laakso, 2009), have long been associated with second language (L2) development (e.g., Kormos, 2000a). To our knowledge, however, no research has looked at the evolution of *self-repair correctness patterns*, that is, the correctness of elements targeted for repair and the correctness of the repair outcomes. Consequently, the present study sought to investigate changes in the self-repair behaviour of English-speaking L2 learners of French over the course of a 5-week period and to verify whether any changes occurred over time. Speech samples of the L2 were collected from 50 adult participants through an elicited narration task at the beginning (Time 1) and the end (Time 2) of a 5-week immersion program. Overall, the results showed that there were qualitative and quantitative changes in self-repairs types, and that correctness of the element being repaired increased significantly over time.

**Résumé**

Les *autoreformulations*, soit toute modification du discours amorcée et réalisée par le locuteur (Salonen et Laakso, 2009), sont depuis longtemps associées au développement des langues secondes (L2) [p. ex. Kormos, 2000a]. Toutefois, à notre connaissance, aucune étude n'a à ce jour examiné l'évolution des *patrons de justesse d'autoreformulations*, à savoir la justesse de l'élément étant l'objet d'une reformulation et du résultat de cette reformulation. Ainsi, la présente étude avait pour objectif d'examiner les autoreformulations d'apprenants anglophones du français L2 pour vérifier si des changements ont eu lieu au cours d'une période de 5 semaines. Les productions langagières de 50 participants ont été obtenues au moyen d'une tâche de narration à partir d'images, au début (temps 1) et à la fin (temps 2) d'un programme d'immersion. De façon générale, les résultats mettent en lumière des changements tant qualitatifs que quantitatifs dans les types d'autoreformulations ainsi qu'une amélioration significative de la justesse des reformulations.

## Evolution of Self-Repair Behaviour in Narration Among Adult Learners of French as a Second Language

### Introduction

A widely noted feature of second language (L2) learners' speech production is the starting and then sudden stopping of a speech segment in an attempt to modify or repair the content of an utterance. These *self-repairs* can be categorized as self-initiated or other-initiated (Schegloff, Jefferson, & Sacks, 1977). *Self-initiated self-repairs*, defined as post-articulatory revisions of speech that speakers themselves initiate and complete (Salonen & Laakso, 2009), are particularly interesting in L2 acquisition research since they denote "an explicit focus on or manipulation of language forms which interrupts the normal flow of production" (Griggs, 1997, p. 407), thus providing information about the cognitive processes in which learners are engaged. Consequently, they have been frequently used as a powerful tool for observing and understanding L2 speech production (e.g., Arroyo, 2003; Camps, 2003; Fathman, 1980; Gilabert, 2007; Griggs, 1998, 2007; Kormos, 1998, 1999a, 1999b, 2000b; Lennon, 1984, 1990, 1994; O'Connor, 1988; Simard, Bergeron, Liu, Nader, & Redmond, 2016; Simard, Fortier, & Zuniga, 2011; Verhoeven, 1989).

Most of the research examining self-initiated self-repairs has been related to developing repair typologies and examining the role of L2 knowledge and the influence of various task types on self-repair distribution frequencies for those typologies. However, to the best of our knowledge, no study has yet examined the evolution of *self-repair correctness patterns*, that is, the linguistic correctness of the elements targeted for repair and the correctness of the ensuing repair outcomes. A better understanding of such correctness patterns would offer insight into L2 linguistic development and contribute to the understanding of L2 speech production development. Therefore, the present study set out to investigate this aspect of L2 self-repair behaviour among adult English-speaking French L2 learners over a 5-week period.

### Self-Repairs in Speech Production

Self-repairs have been conceptualized in research according to Levelt's blueprint modular model (1989, 1999; later adapted for L2 speech production, see de Bot, 1992; Segalowitz, 2010). In this model, the first stage of speech production consists of *conceptualization*, that is, the transformation of speakers' ideas into a preverbal plan. During the first phase of conceptualization, known as *macroplanning*, speakers determine which concepts to include in the emerging utterance and how to spatially and temporally represent them. These ideas are then cast into propositional form during *microplanning*. As they become available, segments of the propositional form of the message are passed on to the *formulator* for morphophonological encoding and articulation. As for self-repairs, they occur at any point in the production process when speakers detect and modify errors or unintended elements in the emerging utterance (e.g., Kormos, 1999b, 2006; Levelt, 1989, 1999).

According to this model, attention[1] plays a key role in speech production, as it must be allocated to each stage of production to detect mismatches between speakers' intentions and production outcomes through self-monitoring processes. The product of this self-monitor corresponds to self-repairs (e.g., Izumi, 2003; Kormos, 1999a, 2006; Levelt, 1989,

1999), which can be triggered in response to the identification of mismatches stemming from all points of the speech production process.

Most repair typologies emerging from studies based on the Levelt model (e.g., Bange & Kern, 1996; Kormos, 1999b) group self-repairs into two general levels of processing: *discourse-level conceptual repairs*, where speakers change the informational structure of the message being conveyed in response to a perceived problem or lack of clarity; and *lower-level formulator repairs*, which occur in response to perceived inaccuracies originating from the morphophonological encoding processes (see Levelt, 1989, 1999).

As for the structure of such repairs, Levelt (1983)[2] identified three essential components: a *reparandum*, that is, the element of the original production targeted for repair; an *editing phase* signaled by the interruption of production, typically in the form of a pause or an editing term such as "eh" or "I mean"; and finally a *reparatum*, which refers to the resulting repair outcomes (Levelt, 1983). The following example, taken from our data, illustrates the three self-repair components, with the reparandum in bold print, the editing phase underlined and the reparatum in italic print.

"L'homme a perdu [**son sac**, <u>ehh</u> *sa valise*]"
"The man lost [**his sack**, <u>ehh</u> *his suitcase*]"

As the speaker detected this conceptual-level lexical mismatch, production was halted at the reparandum ("his sack") and the event was signalled by the editing term ("ehh") while the repair reparatum ("his suitcase") was executed.

**Research on L2 Self Repairs**

Much L2 self-repair research has focused on determining the frequency and distribution of the linguistic elements targeted for repairs (e.g., Bange & Kern, 1996; Fathman, 1980; Gilabert, 2007; Golonka, 2006; Griggs, 1998; Kormos, 1998; Lennon, 1984; Poulisse & Bongaerts, 1994; van Hest, 1996). This repair-typology-generating research has shown, in particular, that L2 speakers do not equally target all stages of the production process for repairs (e.g., Fathman, 1980; Lennon, 1984). For example, L2 speakers tend to produce significantly more lexical repairs than grammatical encoding repairs (Fathman, 1980; Lennon, 1984). They also appear to produce significantly more repairs than first language (L1) speakers (e.g., Bange & Kern, 1996; Poulisse & Bongaerts, 1994; Wiese, 1984; van Hest, 1996) and tend to produce lower-level grammatical or lexical encoding repairs and fewer discourse-level conceptual repairs than native speakers (van Hest, 1996). As language knowledge increases, however, L2 self-repair frequency decreases (Declerck & Kormos, 2012; van Hest, 1996), and the reparanda of those repairs tend to shift from form to discourse-level features (Kormos, 2000a; O'Connor, 1988; van Hest, 1996; Verhoeven, 1989).

To explain the influence of L2 proficiency on self-repair behaviour, researchers, such as Kormos (2000b), O'Connor (1988), van Hest (1996), and Verhoeven (1989), have argued that as lower-level morphophonological formulator processes automatize they become less error-prone and place fewer cognitive demands on the L2 speaker, freeing up attentional resources for discourse-level processing and monitoring. Thus, from a developmental perspective, it would be expected that, as learners' L2 knowledge develops

and automatizes over time, qualitative changes in repair behaviour should be observable, with attention gradually shifting from lower-level grammatical encoding repairs to higher-level discourse repairs.

In light of such an explanation, observable differences over time with regard to the correctness of the reparanda and reparata of the elements targeted for repair might also be expected. For example, Segalowitz's (2010) L2 adaptation of Levelt's (1989, 1999) blueprint of the monolingual speaker posits that as lower-level grammatical encoding processes automatize, learners become less prone to speech production disfluencies, producing fewer grammatical errors, and accordingly fewer incorrect potential reparanda. It might therefore be predicted that the proportion of correct reparanda and reparata (repair outcomes) would increase over time, with a particularly strong increase in correctness for repairs related to the formulator (e.g., verb, determiner, noun and adjective agreements) and the articulator (e.g., pronunciation) since those processes automatize significantly more than those of the conceptualizer (de Bot, 1992).

We refer to the correctness of reparanda and reparata as *correctness patterns*, that is, the correctness of elements targeted for repair (reparandum) and the correctness of the repair outcomes (reparatum). Observation of such patterns provides an additional source of information regarding the participants' focus on language forms. Examining reparanda correctness, for example, might not only offer insight into L2 knowledge development but also into the way in which correctness and language knowledge interact with how L2 speakers allocate attention during production. Observation of reparata correctness does not offer insight into attention allocation, but rather into the way varying types of linguistic reparanda are replaced with correct forms, providing additional information about the production process. To our knowledge, although no prior study has yet investigated the development of the reparanda-reparata correctness patterns in L2 self-initiated self-repairs, one study has examined such features from a cross-sectional perspective.

In their study, Matea Kovač and Milatović (2012) looked at the distribution and correct repair outcomes of self-repairs produced by 101 native speakers of Croatian with different levels of proficiency in English. The participants carried out five oral production tasks, including description and narration tasks. The results of the analyses revealed first that 75% of the repairs targeted errors produced by the participants, while only 8.5% of the repairs targeted discourse-level features resulting in a modification of the ensuing message. The authors interpreted these results as an indication that their participants had a rather low level of proficiency. They argued that "more proficient speakers allocate their attentional resources to enriching the propositional content on the level of discourse, whereas less proficient speakers, due to insufficient knowledge, concentrate on repairing errors arising at lower levels of language processing" (Matea Kovač & Milatović, 2012, p. 239).

The findings of Matea Kovač and Milatović's (2012) study also indicated that a majority of repairs targeted lexical items, while syntactic repairs represented the second largest category. Regarding self-repair outcomes, their results showed that, overall, 82% of the reparata were successful, which, they argued, was due to the larger proportion of successful lexical repairs compared to the smaller proportion of successful syntactic and morphological ones. While these results appear to corroborate previous cross-sectional research (e.g., Kormos, 2000a; O'Connor, 1988; van Hest, 1996; Verhoeven, 1989), they do not offer insight into the development of reparanda-reparata correctness patterns across time.

The present study therefore adds to current knowledge about self-initiated self-repairs by attempting to provide potential answers to the following research question: Are there changes in the reparandum-reparatum correctness patterns of self-initiated self-repairs over time?

## Method

The present study investigated the development of reparandum-reparatum correctness patterns of self-initiated self-repairs (i.e., repairing correctly an incorrect form, incorrectly repairing an incorrect form, etc.) of English-speaking adult learners of French as an L2.

### Participants

Participants for this research were taken from a larger study investigating individual differences and the development of L2 fluency in adult immersion programs. The sample in this study consisted of 50 English-speaking adults (18 males, 32 females) from both the United States and Canada (ranging in age from 18 to 24 years, $M_{age}$ = 23.35 years). They were enrolled in a 5-week French immersion program at a Quebec university where, each week, they received 21 hours of formal instruction and participated in 35 hours of sociocultural activities designed specifically to increase contact with the L2 community. The program curriculum targeted the explicit teaching of some morphosyntactic elements of French (e.g., verb conjugation, plurals, gender); however, a strong version of communicative language teaching was adopted in all classes, placing the pedagogical focus on the development of oral fluency skills. Prior to enrolling in the 5-week program, most (80%) had previously received formal L2 French instruction either in elementary school or high school, or both. Participants' knowledge of French was assessed using the Laval Placement Test (Centre international de recherche sur le bilinguisme, 1976),[3] which, according to test specifications, showed learners to be at a low-intermediate level at the onset of the study ($M$ = 51.48, $SD$ = 15.34, Min. = 47.11, Max. = 57.87).

All participants signed a language pledge committing to speaking only French during the 5-week period, and the majority (35 out of 50) also chose to live with a French-speaking host family during their stay.

### Instruments

Two measurement instruments were used: a narrative task and a background questionnaire.

**Narrative task.** To gather self-repair data, a picture narrative task was used, as this type of task has been repeatedly shown to elicit reliable samples of oral production in a variety of different contexts (Rossiter, Derwing, & Jones, 2008). It has also been used in previous research examining repair phenomena (Bange & Kern, 1996; Camps, 2003; Gilabert, 2007; Simard et al., 2011; Simard et al., 2016), and, more importantly, it is also suggested that such a task promotes learners' use of repairs in oral production (e.g., Gilabert, 2007).

The picture narrative selected for this study was the Suitcase Story (Derwing, Munro, Thomson, & Rossiter, 2009), which has been used in numerous L2 speech production studies (e.g., Derwing et al., 2009; Derwing, Rossiter, Munro, & Thompson, 2004). The story consists of an 8-frame storyline in which a man and woman carrying identical green suitcases while walking toward the corner of a busy city street accidentally bump into one another, causing them to fall to the ground. After standing back up, they mistakenly retrieve the wrong suitcase and do not notice the error until they arrive at their respective hotel rooms.

**Background Questionnaire.** A background questionnaire provided information about participants' age, sex, L1, level of education, and amount of previous L2 instruction in French and other languages. The information gathered from the questionnaire was used to develop a detailed socio-demographic profile of the learners participating in the program. Given the objectives of the present study, only information on age, sex, L1 background, and previous L2 instruction are reported.

## Procedure

All participants were met individually and informed that they would have to recount a story aloud through the use of picture cues, take a language knowledge test, and complete a background questionnaire. They were then invited to look at the 8-picture storyline and to mentally prepare for their narration, but were not allowed to take notes or ask clarification questions about lexical content related to the picture frames. The participants were allowed 1 minute to plan their story before recording. Indeed, evidence has shown that pre-task planning lightens the cognitive load of production tasks, resulting in an increase in the fluency (e.g., Foster & Skehan, 1996) and the complexity of output (e.g., Crookes, 1989; Ellis, 1987; Foster & Skehan, 1996; Ortega, 1999; Yuan & Ellis, 2003), thereby increasing self-repair frequency. After the 1-minute preparation period, participants were then instructed to tell their version of the story within 3 minutes. The picture-cue task was administered at the beginning, Time 1 (T1), and end, Time 2 (T2), of the 5-week study. The language knowledge task and background questionnaire were given once at T1.

We are aware of the potential test-retest effects that may arise when using the same narrative task at T1 and T2; however, in the present study context, it was particularly important to ensure that self-repair behaviour was prompted under similar task conditions in order to make more reliable comparisons across time with respect to the development of specific repair categories (see below for categories). Additionally, it can be argued that the 5-week window between the two measures also minimizes the possible learning effect caused by the test-retest (see Wilson & Putman, 1982). Consequently, it was decided that the potential shortcomings of using the same speech task twice were minimized by the importance of obtaining comparable speech samples at both testing sessions and by the 5-week lapse between the two.

## Data Coding

The stories were digitally recorded, and the resulting recordings were orthographically transcribed and then blind-coded for repairs by three judges. Following procedures in Levelt (1989), the judges identified 362 repairs in the 12,445-word corpus by

identifying the reparandum and the reparatum. The inter-rater reliability was then calculated across three sets of observations (Cronbach's Alpha = .972). Afterward, the three judges pinpointed all discrepancies in their coding and attempted to resolve these via discussion. In situations where no agreement could be reached for a given repair, it was removed from the corpus. Following this procedure, a total of 19 repairs were removed from the corpus.

After establishing the initial corpus of repairs, the judges further coded the repairs into two subcategories. In this regard, Simard et al. (2011) pointed out that while self-repair categories vary from one study to another and are operationalized in different ways (e.g. Bange & Kern, 1996; Kormos, 1999a), one distinction found in all self-repair studies is between what are referred to as form repairs (F-Repair) and choice repairs (C-Repairs). *F-repairs* refer to modifications to linguistic form while *C-Repairs* signal modifications changing the conceptual content or meaning of utterances. The judges therefore coded the 343 remaining repairs as either F-Repairs ($n = 222$) or C-Repairs ($n = 121$) [see Simard et al., 2011].

The F-Repairs were further coded into subcategories according to the targeted formal aspects: gender ($n = 61$), number ($n = 30$), conjugation ($n = 90$), phonology ($n = 34$) and syntax ($n = 7$). Table 1 presents examples and explanations of the various F-Repair types.

Table 1
*Examples of the F-Repair Types*

| F-Repair Type | Example | Repair Target |
|---|---|---|
| Gender | "Il va **à le,** *à la* maison . . ." (He goes **to the,** *to the house* . . .) | Determiner gender agreement |
| Number | "J'espère **que le,** *que les gens* . . ." (I hope **that the,** *that the people* . . .) | Determiner number agreement |
| Conjugation | "Quand il **va arrive,** *va arriver* . . ." (When he **going comes,** *is going to come. . .*) | Verbal morphology |
| Phonology | "Il prend **le levli,** *la valise* . . ." (He takes **the sluit,** *the suitcase* . . .) | Pronunciation |
| Syntax | "**J'ai lui,** *je lui ai demandé.* (I **him asked,** *asked him*.) | Pronoun placement |

*Note.* The reparandum is indicated in bold print and the reparatum in italics.

In all of these examples, the speaker repaired some aspect of linguistic form. With the gender repair, for example, the speaker detected the erroneous masculine determiner *le* (reparandum) and replaced it with the correct feminine form *la* (reparatum).

*C-Repairs*, which imply conceptual-level changes to the emerging message, were also divided into subcategories according to whether they targeted a specific word ($n = 83$) or a determiner ($n = 38$). Table 2 provides examples of such repairs.

Table 2
*Examples of C-Repair Types*

| F-Repair Type | Example | Repair Target |
|---|---|---|
| Word | " **le chien,** eh *le garçon* est parti . . ." <br> (and then **the dog**, eh *the boy* ran off . . .) | Replacement of the word *dog* with *boy* |
| Determiner | "**à la**, *à une personne*" <br> (**to the**, *to a person*) | Replacement of definite with indefinite determiner |

These C-Repairs entail the replacement of one linguistic element with another. For the word repair, the speaker replaced the erroneous noun group *le chien* (reparandum) with *le garcon* (reparatum)*,* and for the determiner repair, the speaker replaced an indefinite article with a definite one without changing the gender or number.

For each participant, the aggregate repair rate per number of words was calculated according to the method used by Griggs (1997), where "only words conveying new information were included, and therefore second parts of repeats, false starts and repeated and reformulated words in repair sequences were discounted, as were meta-comments in French" and English (p. 410). Then, the ratios were calculated by dividing the raw number of reformulations produced in three minutes (R) by the number of words produced during the narration (W) multiplied by 100:

$$\left( \frac{R}{W} \right) \times 100 = Ratio$$

Finally, the correctness patterns of all attempted repairs were coded using the following categories:

1. Incorrect-Correct, where an incorrect reparandum is repaired with a correct reparatum;
2. Incorrect-Incorrect, where an incorrect reparandum is repaired with an incorrect reparatum;
3. Correct-Correct, where a correct reparandum is repaired with a correct reparatum;
4. Correct-Incorrect, where a correct reparandum is repaired with an incorrect reparatum.

The coded correctness patterns were assessed independently by two judges and the interrater reliability coefficient was .986, suggesting an exceptionally high agreement between judges with respect to the four accuracy categories.

**Data Analyses**

First, we verified the normality of the distribution of our data using the Shapiro-Wilk test and a calculation of skewness ratios (Larson-Hall, 2010). Then, we examined the overall self-initiated self-repair production. In order to determine the significance of T1-T2

change, we calculated the percentage of self-repair change occurring at the end of the 5-week program (T2). Additionally, frequencies were calculated to obtain information about T1-T2 changes in correctness patterns for specific language features. The magnitude of changes in reparatum correctness patterns was tracked through the proportion of correct repair outcomes from T1 to T2. We compared the ratios of repairs and repair patterns obtained at T1 and T2 using a Wilcoxon procedure. We then reported effect size for the Wilcoxon using a percentage variance measure (*r*), as recommended by Larson-Hall (2010), and interpreted the scores obtained as small effect size for .1, medium effect size for .3 and large effect size for .5 and above. The confidence interval was set at 95%.

## Results

We first present the overall self-initiated self-repair results and then expose the correctness patterns.

### Self-Initiated Self-Repairs Overall Results

First, we verified the normality of the distribution of the total self-repairs generated at T1 and T2. Table 3 reports the skewness and kurtosis values as well as the result obtained from the Shapiro-Wilk.

Table 3
*Distribution of Results: Time 1 and Time 2 Self-Initiated Self-Repairs*

|  | Skewness $SE = .377$ | Kurtosis $SE = .662$ | Shapiro-Wilk |
|---|---|---|---|
| T1 | 1.252 | .940 | .851  $p < .001$ |
| T2 | 1.438 | 2.844 | .865  $p < .001$ |

*Note. n* = 50; T1 = Time 1; T2 = Time 2.

The results in Table 3 indicate a sharp positive skew for both T1 and T2 repairs that could not be corrected through log transformations.

Table 4 reports the total number of repairs and the means and standard deviations for self-repair ratios for each repair category at T1 and T2. The last column of the table shows the percentage of repair ratio change from T1 to T2.

Table 4
*Time 1 and Time 2 Repair-Type Distribution and Difference*

| Repairs Types | T1 | | | T2 | | | T1-T2 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | Total Repairs | *M* Ratio | *SD* | Total Repairs | *M* Ratio | *SD* | Ratio % Change |
| Total Repairs | 192 | 3.55 | 2.09 | 151 | 2.15 | 2.04 | -39** |
| F-Repairs | 122 | 2.26 | 1.77 | 100 | 1.42 | 1.69 | -37** |
|    Gender | 44 | .80 | 1.30 | 17 | .24 | .49 | -70** |
|    Number | 14 | .25 | .59 | 16 | .23 | .49 | -8 |
|    Conjugation | 42 | .78 | 1.08 | 48 | .69 | .84 | -12 |
|    Phonological | 17 | .32 | .64 | 17 | .24 | .47 | -25 |
|    Syntax | 5 | .10 | .45 | 2 | .03 | .13 | -70 |
| C-Repairs | 70 | 1.30 | 1.39 | 51 | .72 | .85 | -45 |
|    Word | 45 | .82 | 1.12 | 38 | .55 | .73 | -33 |
|    Determiner | 25 | .47 | .81 | 13 | .18 | .45 | -61 |

*Note.* $n = 50$; T1 = Time 1; T2 = Time 2.
*$p < .05$. **$p < .01$.

The data in Table 4 first indicate that our participants made on average 3.55 self-repairs per 100 words of discourse in T1. After the 5-week immersion program, however, the average repairs for the same group dropped to 2.15 self-repairs per 100 words. Among the T1 repairs, there were about 75% more F-Repairs (2.26) than C-Repairs (1.30), while in T2, C-Repairs (.72) appeared to drop more drastically than F-Repairs (1.42). Among the F-Repairs, gender (.80) and conjugation (.78) repairs were the most frequent type, which were followed by phonological (.32) and number (.25) repairs. Finally, syntax (.10) repairs were very infrequent. With regard to the C-Repairs, at T1, there were about 75% more word repairs (.82) than determiner repairs (.47). At T2, determiner repairs (.18) decreased markedly by almost 60%, virtually doubling the observed drop of about 33% for word repairs (.55). Overall, this pattern of results suggests that both the nature and frequency of learners' repairs appeared to change over the 5-week study period, pointing to a reduction in repair frequency.

Overall, a Wilcoxon procedure conducted on the data revealed a significant shift in repair behaviour at the end of the program. The 39% drop in total repairs was indeed significant ($Z = -3.934$, $p < .001$, $r = .55$) with a large effect size. With respect to the two subcategories, there was a significant 37% reduction in F-repairs ($Z = -2.907$, $p = .003$, $r = .41$). The F-repair decrease can clearly be attributed to a sharp reduction of 70% in gender repairs ($Z = -3.211$, $p < .001$, $r = .45$). The decline in C-Repairs was also significant ($Z = -1.937$, $p = .05$, $r = .27$), but with a small effect size. The tests revealed that, among C-Repairs, only the drop in determiner repairs was significance ($Z = -2.058$, $p = .04$, $r = .29$). Finally, differences between F-Repairs and C-Repairs were significant at both pretest ($Z = -2.785$, $p = .005$, $r = .39$) and posttest ($Z = -2.730$, $p = .006$, $r = .38$).

**Reparandum-Reparatum Correctness Patterns**

Next, we looked more specifically at the reparandum-reparatum correctness patterns in our participants' self-repairs. The results of the analyses verifying the normality of the distribution for those patterns are presented in Table 5.

Table 5
*Distribution of Results: Time 1 and Time 2 Reparandum-Reparatum Correctness Patterns*

|    |                    | Skewness  *SE* = .377 | Kurtosis  *SE* = .662 | Shapiro-Wilk | |
|----|--------------------|----------|----------|------|-----------|
| T1 | Incorrect-Correct   | 1.48 | 2.86  | .866 | *p* < .001 |
|    | Incorrect-Incorrect | 1.11 | .429  | .765 | *p* < .001 |
|    | Correct-Correct     | 3.12 | 1.44  | .344 | *p* < .001 |
|    | Correct-Incorrect   | 4.35 | 22.60 | .410 | *p* < .001 |
| T2 | Incorrect-Correct   | 2.25 | 7.85  | .783 | *p* < .001 |
|    | Incorrect-Incorrect | 1.95 | 2.87  | .584 | *p* < .001 |
|    | Correct-Correct     | 3.36 | 13.18 | .493 | *p* < .001 |
|    | Correct-Incorrect   | 2.50 | 5.66  | .487 | *p* < .001 |

*Note. n* = 50; T1 = Time 1; T2 = Time 2.

Table 5 shows that, in line with the general repair data reported in Table 3, the correctness pattern results were not normally distributed.

Table 6 reports the means and standard deviations for the T1 and T2 correctness patterns. Wilcoxon tests were used to determine whether the T1 and T2 differences in these patterns were statistically significant.

Table 6
*Means and Standard Deviations for Reparandum-Reparatum Correctness Patterns and Time 1-Time 2 Difference*

| Accuracy Patterns | T1 | | T2 | | Difference | % of Change |
|-------------------|------|------|------|------|------|------|
|                   | *M* | *SD* | *M* | *SD* | | |
| Incorrect-Correct   | 2.30 | 1.78 | 1.45 | 1.62 | -.85 | -37** |
| Incorrect-Incorrect | .66  | .88  | .27  | .52  | -.39 | -59*  |
| Correct-Correct     | .10  | .31  | .30  | .71  | .20  | 200*  |
| Correct-Incorrect   | .33  | .91  | .15  | .36  | -.18 | -54   |

*Note. n* = 50; T1 = Time 1; T2 = Time 2.
*\*p* < .05. \*\**p* < .01.

Inspection of Table 6 shows that the Incorrect-Correct pattern (i.e., incorrect reparandum, correct reparatum) was the most common at both T1 (2.3) and T2 (1.45), accounting for 67% of repairs in both times. In other words, most repairs at T1 and T2 targeted an incorrect reparandum and resulted in a correct reparatum. The Incorrect-Incorrect pattern, that is, instances where both the reparandum and reparatum were incorrect, was the second most common correctness pattern at T1, constituting nearly a third of all repairs (.66). This type of pattern decreased precipitously at T2, accounting for only 12% of the total repairs (.27). Correct-Correct repairs, where both the reparandum and reparatum were correct, were the least common at T1, making up only 3% of repairs (.10). This pattern, however, was the only one that increased from T1 to T2, finishing at 13% (.30). Finally, Correct-Incorrect repairs, where speakers rendered incorrect reparatum to a correct reparandum, constituted almost 10% of all repairs at T1 (.33) and decreased to a meager 7% at T2 (.15).

Table 6 also shows that the 37% drop in Incorrect-Correct repairs, the most common of correctness patterns, was statistically significant ($Z = -2.805$, $p = .005$, $r = .39$). With regard to the second most common type, the 59% drop in Incorrect-Incorrect repairs also reached significance ($Z = -2.475$, $p = .01$, $r = .35$). Furthermore, the ratio of Correct-Correct repairs doubled over time and this increase was significant ($Z = -2.166$, $p = .03$, $r = .30$), while the 54% drop in Correct-Incorrect repairs was not.

Changes in the proportion of correct repair outcomes from T1 to T2 are presented in Table 7.

Table 7
*Proportion in Percentage of Correct Reparata at Time 1 and Time 2*

|  | T1 | T2 | Difference | % of Change |
|---|---|---|---|---|
|  | M | M |  |  |
|  | (SD) | (SD) |  |  |
|  | 71.04 | 85.04 | 14 | 20* |
|  | (5.10) | (3.32) |  |  |

*Note.* $n = 50$; T1 = Time 1; T2 = Time 2.
*$p < .05$. **$p < .01$.

Table 7 shows that among the learners ($n = 50$), accurate reparata outcomes increased from 71% at T1 to 85% at T2, which constituted a significant 20% increase ($Z = -2.364$, $p = .02$, $r = .33$). Also noteworthy was the decrease in the standard deviation from T1 (5.1) to T2 (3.3), pointing to a reduction of variance in repair behaviour among participants across the immersion experience.

Finally, changes in correctness patterns for specific language features are reported in percentages in Table 8.

Table 8
*Percentages (%) of Language Aspects per Correctness Patterns at Time 1 and Time 2*

|  | Incorrect-Correct | | Incorrect-Incorrect | | Correct-Correct | | Correct-Incorrect | |
|---|---|---|---|---|---|---|---|---|
|  | T1 | T2 | T1 | T2 | T1 | T2 | T1 | T2 |
| F-Repairs | 67 | 71 | 49 | 65 | 71 | 72 | 62 | 71 |
| Gender | 24 | 9 | 0 | 6 | 0 | 12 | 23 | 29 |
| Number | 9 | 10 | 3 | 0 | 14 | 32 | 15 | 14 |
| Conj. | 15 | 33 | 37 | 53 | 43 | 28 | 23 | 29 |
| Phono. | 15 | 15 | 3 | 6 | 14 | 0 | 0 | 0 |
| Syntax | 3 | 4 | 6 | 0 | 0 | 0 | 0 | 0 |
| C-Repairs | 33 | 29 | 51 | 35 | 29 | 28 | 38 | 29 |
| Word | 25 | 20 | 31 | 24 | 0 | 12 | 15 | 29 |
| Det. | 9 | 9 | 20 | 12 | 29 | 16 | 23 | 0 |

*Note.* T1 = Time 1; T2 = Time 2.

The frequency distribution presented in Table 8 first shows that considerably more Incorrect-Correct repairs targeted F-Repairs (67%) than C-Repairs (33%). Incorrect-Incorrect repairs, however, were roughly divided between F-Repairs (49%) and C-Repairs (51%) at T1 but shifted toward F-Repairs (65%) at T2. Self-repairs with a Correct-Correct

accuracy pattern increased at T2 for both F-Repairs and C-Repairs, but the C- and F-Repair proportions remained mostly unchanged. Finally, for the Correct-Incorrect pattern, only very low frequencies were observed, and the proportions reveal a shift toward more F-Repairs and fewer C-Repairs.

With regard to the correctness pattern subcategories, on one hand there was an important decrease from T1 to T2 in the Incorrect-Correct pattern for gender (62.5%, F-Repairs) and for words (20%, C-Repairs). On the other hand, there was only one notable increase in proportions from T1 (15%) to T2 (33%) for the Incorrect-Correct accuracy pattern observed for repairs targeting morphological aspects of the participants' speech production (conjugation). For the Incorrect-Incorrect accuracy patterns, we observed a 16-point increase in conjugation F-Repairs and a proportionate 16-point drop in C-Repairs targeting words (7-point decrease) and determinants (9-point decrease). For Correct-Correct, there was an increase in frequencies for all categories, and while the F- and C-Repair proportions remained unchanged, the subcategories proportions shifted in both directions: While there was a sharp decrease in the proportion of conjugation (-15), phonological (-14), and determiner (-13) repairs, there was an increase in gender (+12), number (+18), and word repairs (+12). The low frequency of Correct-Correct repairs (3% in T1 and 13% in T2), however, does not allow us to make claims about these subcategory shifts with regard to L2 development.

## Discussion

This study examined whether there were changes in the reparandum-reparatum correctness patterns of self-initiated self-repairs over time by assessing the self-initiated self-repairs of 50 English-speaking L2 French adult learners at two points in time over the course of a 5-week study period.

We first verified whether there were changes in self-repair accuracy patterns by examining the ratios of self-repairs. As expected, our results revealed a significant reduction in repairs ($p < .001$), decreasing from a ratio of 3.5 repairs per 100 words at T1 to 2.15 repairs per 100 words at T2. Such a drop in repair ratios can certainly be explained by an increase in production across time, with participants producing an average of 109 words at T1 and 140 words at T2, which can be seen as an indication of increased French language proficiency.

In the present context, the observed decrease in repair frequency appears to be more substantial than the modest differences previously observed in van Hest (1996), where beginner and intermediate learners made 2.65 and 2.55 repairs per 100 words respectively. The differences with regard to the magnitude of the repair frequency decrease may be largely due to the fact that the present study is longitudinal while the van Hest study was cross-sectional, and therefore was unable to further track changes in self-repair behaviour. Additionally, van Hest investigated the self-repairs of beginner and intermediate teenagers (adolescents). In the present study, however, beginner and intermediate adult self-repair behaviours were examined; therefore, it might be that our adult learners were much more oriented toward accuracy and thus made fewer overall self-repair attempts than the adolescent learners reported in van Hest. Finally, the type of instruction might have also impacted self-repair frequencies. For example, in our study, instruction focused primarily on the morphosyntactic aspects of French such as conjugation and gender agreement, which may very well have contributed to the greater reduction in F-repairs observed at the end of

the study. However, since van Hest did not detail the instructional treatment used with the learners in his study, it is of course difficult to know to what extent explicit instruction of French morphology in the present context influenced the overall decrease in F-repairs.

With regard to repair types, we observed a reduction in both F-Repairs (37%, $p < .01$) and C-Repairs (45%, $p < .05$). The effect size for the C-Repairs was, however, quite small, suggesting that the immersion experience may have had the greatest impact on improving learners' formulator processes, in particular, gender agreement assignment. These findings are in line with previous empirical studies (e.g., Matea Kovač & Milatović, 2012; O'Connor, 1988; van Hest, 1996) revealing reductions in grammatical encoding repairs as proficiency increases. These findings also support Levelt-inspired L2 speech production models (e.g., de Bot, 1992; Segalowitz, 2010) which depict, from a developmental perspective, conceptualizer processes as being relatively stable and universal, and formulator processes as comparatively unstable and evolving as learners gain in proficiency. That is, while conceptualizer processes can never fully automatize, because speakers always need to pay some attention to their communicative intentions and the evolving discourse model, unstable low-proficiency formulator processes gradually stabilize and automatize to the point where they largely operate outside of conscious awareness with L2 development. This stabilization results in fewer disfluencies and ensuing repairs (Segalowitz, 2010). The significant reduction in F-Repairs we observed in this study seems to support this claim.

Regarding correctness patterns, our results revealed, as predicted, an overall increase in accuracy of both the reparandum (repair targets) and the reparatum (repair outcomes) of self-repairs. The proportion of repairs targeting erroneous linguistic reparanda dropped by 16% while those repairs with linguistically accurate targets increased by 15%. According to Segalowitz's (2010) model, this increase in the accuracy of the elements targeted for repair reflects developing and stabilizing formulator processes that generate fewer and fewer disfluencies and repairs. This in turn frees up more attentional resources for the monitoring of discourse level features for ambiguities, which are often linguistically correct (Kormos, 2000a). Paralleling the reparanda accuracy trends, our analyses also reveal a significant increase in reparata accuracy, which jumped from 71% in T1 to 85% in T2 (see Table 7). First we remark that this reparata accuracy reflects findings from previous L2 self-repair studies (e.g., Kormos, 1998; Matea Kovač & Milatović, 2012) observing about an 80% accuracy rate for repair outcomes. What is interesting about these data, however, is the 14-point increase in correct reparata. Inspection of the subcategories suggests that this increase can be attributed to greater accuracy in conjugation, gender, and number F-Repairs, and word C-Repairs.

Among the accuracy pattern repair types, the Correct-Correct repairs category was the only one that increased significantly from T1 to T2 (200%). Such a change, reflecting the drop in F-Repairs discussed above, is again most likely due to L2 development and a stabilizing formulator, which frees up attentional resources for the monitoring of discourse-level features (Kormos, 2000a). Indeed, by nature, Correct-Correct repairs entail modifications to the conceptual structure of emerging utterances (i.e., word and determiner choice) rather than to the repair of morphophonological errors. However, following this logic we would have also expected an increase in C-Repairs, which was not observed. This can be explained by the relatively small proportion of Correct-Correct repairs in the corpus, which increased from 3% to 13% at the end of immersion. This increase was not sufficient enough to contribute to a significant change in the overall proportion of C-Repairs.

Finally, similar to Matea Kovač and Milatović (2012), in our study, the most successful self-repairs targeted forms. We would argue that this is due to the emphasis that is put on the teaching of certain morphosyntactic aspects in the French language program in which our participants were enrolled. Such emphasis during instruction would not only increase exposure to given morphosyntactic features, but also multiply opportunities for explicit feedback and self-repair. Indeed, Griggs (1997) showed that frequent self-repairs facilitate L2 linguistic development, because the targeted structures benefit from the enhanced allocation of attentional resources essential for acquisition. The initial increase in self-repairs of a given formal aspect, stimulated by instruction, would appear to facilitate its development, and lead to a subsequent decrease in the targeting of those reparanda for self-repair and an increase in the proportion of correct reparanda outcomes for those remaining repairs. The results of the present study only point in this direction, they do not allow us to make links between instruction, self-repair correctness, and L2 development. Further research operationalizing these three variables would be necessary to draw such conclusions.

## Conclusion

Our study set out to examine longitudinal changes in the frequency of self-initiated self-repair behaviour and the correctness patterns of L2 speakers. We recruited 50 native English speakers participating in a 5-week French L2 language immersion program, in which the classes could be characterized as balanced between meaning- and form-focused instruction. Significant changes in self-repair behaviour were observed. The repair data showed a significant reduction in self-repair frequency, with stronger effect sizes for F-repairs. The correctness patterns also shifted, revealing a reduction in repairs targeting incorrect reparanda and an increase in those targeting correct reparanda. There was also a significant increase in the correctness of repairs outcomes. Our main findings suggest that (a) repair behaviour can change in a relatively short period time and that (b) this change appears to occur particularly with respect to F-repairs. The results point to a contrast between the relative instability of formulator processes that proceduralize with L2 practice and the relative stability of conceptualizer processes.

It appears that explicit teaching coupled with the communicative approach learning context in which our participants learned French may have been a catalyst for the observed behavioural changes to occur, suggesting that developmental trajectory of self-repairs is somewhat sensitive to instruction. However, this needs to be further investigated. More specifically, future research should investigate learning contexts with and without explicit instruction in order to determine whether naturalistic exposure alone can impact repair behaviour differently from what we have observed. Furthermore, Griggs (1997) showed that speakers who self-repair more frequently demonstrate L2 development, which suggests that future research should also examine the promotion of self-repair teaching in the classroom as a means to improve students' self-monitoring system for grammar. Finally, in the present study, only a narration task was used to elicit repair behaviour; however, Gilabert (2007) found that task types actually interact with self-repair behaviours. Consequently, it would be important to investigate self-initiated self-repairs in a context using multiple tasks. Findings from such studies would have practical implications for instructional practices.

Correspondence should be addressed to Daphnée Simard.
Email: simard.daphnee@uqam.ca

**Notes**

[1]It should be noted that Levelt (1989, 1999) did not provide information as to the exact nature and contribution of attention to self-repair behaviours. However, as noted in Simard et al. (2011)'s conclusions and further confirmed by Zuniga (2015) and Simard et al. (2016), *attention shift,* that is, the shift of attentional resource allocation to multiple parallel stimuli, seems to be the attentional function involved in self-repair behaviours.

[2]Self-repairs can be either overt or covert (Levelt, 1983). While *overt self-repairs* occur after the articulation of speech, *covert self-repairs* occur before the articulation of speech. As Postma (2000) explained, "covert repairs are often ambiguous and difficult to classify (e.g., does a long pause signal a real covert repair, whereas a short pause does not?)" (p. 106). The concrete nature of overt self-repairs makes them easier to study.

[3]This multiple-choice test targets three specific areas of knowledge in French: (a) the ability to distinguish sounds in spoken French (phonetics), (b) the ability to identify correct elements of the French sentence and its grammatical structure (grammar), and (c) the range of one's French lexical knowledge (vocabulary). It is primarily designed as an instrument used to place students in appropriate-level language courses at the start of the immersion program in which the participants were enrolled. The test yields a maximum score of 100 (phonetics: 16, grammar: 54, vocabulary: 30).

**References**

Arroyo, E. (2003). La reformulation en communication exolingue chez des locuteurs hispanophones parlant français. *Marges linguistiques, septembre-octobre*. Retrieved from http://www.marges-linguistiques.com

Bange, P., & Kern, S. (1996). La régulation du discours en L1 et en L2. *Études romanes*, *35,* 69-103.

Camps, J. (2003). The analysis of oral self-correction as a window into the development of past time reference in Spanish. *Foreign Language Annals*, *36,* 233-242.

Centre international de recherche sur le bilinguisme. (1976). *Laval test, placement test, French as a second language.* Quebec, Canada: Presses de l'Université Laval.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*, 367-383.

de Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, *13,* 1-24.

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition, 15,* 782-796.

Derwing, T. M., Munro, M. M., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31,* 533-557.

Derwing, T. M., Rossiter, M. J., Munro, M. M., & Thomson, R. I. (2004). L2 fluency: Judgments on different tasks. *Language Learning, 54,* 655-679.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition, 9,* 12-20.

Fathman, A. (1980). Repetition and correction as an indication of speech planning and execution processes among second language learners. In H. Dechert & M. Raupach (Eds.), *Toward a cross-linguistic assessment of speech production* (pp. 77-85). Frankfurt, Germany: Verlag Peter D. Lang.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18,* 299-323.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Journal of Applied Linguistics*, *45,* 215-240.

Golonka, E. (2006). Predictors revised: Linguistic knowledge and metalinguistic awareness in second language gain in Russian. *Modern Language Journal, 90,* 496-505.

Griggs, P. (1997). Metalinguistic work and the development of language use in communicative pair work activities involving second language learners. In L. Diaz & C. Pérez (Eds.), *Views on the acquisition and the use of second languages* (pp. 403-415). Barcelona, Spain: Universitat Pompei Fabrat.

Griggs, P. (1998). L'effet de la tâche sur la régulation de l'activité métalinguistique entre apprenants partageant la même langue maternelle. In M. Souchon (Ed.), *Pratiques discursives et acquisition des langues étrangères* (pp. 103-115). Besançon, France: Université de Franchecompté.

Griggs*, P. (2007). Perspective sociocognitive sur l'apprentissage des langues étrangères : situation naturelle ou guidée*. Paris, France: l'Harmattan.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale for the output hypothesis. *Applied Linguistics, 24*, 168-196.

Kormos, J. (1998). A new psycholinguistic taxonomy of self-repairs in L2: A qualitative analysis with retrospection. *Even Yearbook, ELITE SEAS Working Papers in Linguistics, 3*, 43-68.

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

Kormos, J. (1999b). Monitoring and self-repair in L2. *Language Learning*, *49,* 303-342.

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

Kormos, J. (2000b). The timing of self-repairs in second language speech production. *Studies in Second Language Learning, 22,* 145-167.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.

Lennon, P. (1984). Retelling a story in English as a second language. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 50-68). Tübingen, Germany: Narr.

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

Lennon, P. (1994). Self-correction and error in advanced learner spoken narrative. The dynamics of language processes. In G. Bartelt (Ed.), *Essays in honour of Hans W. Dechert* (pp. 85-103). Tübingen, Germany: Gunter Narr Verlag.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, United Kingdom: MIT Press.

Levelt, W. J. M. (1999). Language production: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 83-122). Oxford, United Kingdom: Oxford University Press.

Matea Kovač, M., & Milatović, B. (2012). Analysis of repair distribution, error correction rates and repair correctness in L2. *Studia Linguistica: Journal of General Linguistics, 67*, 225-256.

O'Connor, N. (1988). Repairs as an indicative of interlanguage variation and change. In T. J. Walsh (Ed.), *Georgetown University Round Table in Languages and Linguistics 1988: Synchronic and diachronic approaches to linguistic variation and change* (pp. 251-259). Washington, DC: Georgetown University Press.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21*,109-148.

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77,* 97-131.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*, 36-57.

Rossiter, M. J., Derwing, T. M., & Jones, V. M. L. O. (2008). Is a picture worth a thousand words? *TESOL Quarterly, 42,* 325-329.

Salonen, T., & Laakso, M. (2009). Self-repair of speech by four-year-old Finnish children. *Journal of Child Language, 36*, 855-882.

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

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361-382.

Simard, D., Bergeron, A., Liu, Y.-G., Nader, M., & Redmond, L. (2016). Production d'autoreformulations autoamorcées en langue seconde : rôle de l'attention et de la mémoire phonologique. *Revue canadienne des langues vivantes/Canadian Modern Language review, 72,* 183-210.

Simard, D., Fortier, V., & Zuniga, M. (2011). Attention et production d'autoreformulations autoamorcées en français langue seconde, quelle relation ? *Journal of French Language Studies, 21,* 417-436.

van Hest, E. (1996). *Self-repair in L1 and L2 production*. Tilburg, Netherlands: Tilburg University Press.

Verhoeven, L. T. (1989). Monitoring in children's second language speech. *Second Language Research, 5*, 141-155.

Wiese, R. (1984). Language production in foreign and native languages: Same or different? In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 11-25). Tübingen, Germany: Gunter Narr Verlag.

Wilson, V. L., & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal, 19*, 249-258.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*, 1-27.

Zuniga, M. (2015). *The role of attention in L2 speech production* (Unpublished doctoral dissertation). Université du Québec à Montréal, Montreal, QC.