L2 and L1 repairs
Speech production in a comparative perspective

Susanne Ullenius
L2 and L1 repairs
Speech production in a comparative perspective
Susanne Ullenius

Abstract
I investigated and compared L2 and L1 speech errors and repairs. A speech error may be defined as a linguistic item that is partially or wholly articulated but disagrees with the speaker’s desired communicative intention. A self-repair usually comprises a speech error, a self-interruption, and a repair. Repairs reveal information about the speech production process and in particular about the monitoring component. Errors and repairs were collected from 24 L1 and L2 English speakers who were audio recorded while describing patterns of multi-coloured interconnected nodes. The methodology is a modified version of Levelt’s (1982; 1983) methodology in his study of L1 Dutch speakers, and his results are incorporated in the analysis section for comparison purposes. The hypothesis that L2 speakers produce more repairs than L1 speakers was confirmed. The hypothesis that they produce more lexical errors and less appropriateness errors compared to L1 speakers was confirmed in relation to the English L1 group but not in relation to Levelt’s Dutch L1 group. The hypothesis that L2 speakers leave a larger proportion of their lexical errors unrepairob was not confirmed. The significant differences in numbers and types of errors between the L1 and the L2 data may be related to Paradis’s (2009) theory of declarative knowledge and procedural competence, which entails a higher demand on attentional resources during L2 production. Data may be influenced by methodological inconsistencies, and may also be too small to generalise upon.

Keywords
Contents

1. Introduction ............................................................................................................. 1

2. Background ........................................................................................................... 2
   2.1 An L1 speech production model ........................................................................ 2
   2.2 L2 speech production, automaticity, and attention .......................................... 4
   2.3 Levelt’s study of L1 speech errors and repairs .................................................. 5
      2.3.1 Types of errors and repairs ....................................................................... 5
      2.3.2 The structure of repair .............................................................................. 8
   2.4 Previous studies on L2 monitoring ................................................................... 9

3. Methodology ......................................................................................................... 10
   3.1 Material and procedure ................................................................................... 10
   3.2 Interpreting the data ....................................................................................... 11

4. Results and analysis ............................................................................................. 12
   4.1 Numbers of repairs ......................................................................................... 12
   4.2 Types of repairs .............................................................................................. 13
   4.3 The relation to L2 proficiency ......................................................................... 16

5. Discussion ............................................................................................................ 17
   5.1 Limitations of the study .................................................................................. 17
   5.2 General discussion ......................................................................................... 17

6. Conclusion ............................................................................................................ 18

References ............................................................................................................... 19
Appendix A ............................................................................................................... 20
Appendix B ............................................................................................................... 21
Appendix C ............................................................................................................... 22
Appendix D ............................................................................................................... 23
1. Introduction

In the recent 25 years, the interest for speech production has grown within the fields of both cognitive science and linguistics; a range of serial and interactive theories attempt to explain how an intention is turned into an utterance. Research is often based on speech errors, or slips-of-the-tongue, which indicate how lexical information is stored, accessed, and assembled. A speech error may be defined as a linguistic item that is partially or wholly articulated but disagrees with the speaker’s desired communicative intention (note that it may still be grammatically and semantically suitable in the given context). If the speaker notices the error, she normally makes a repair in order to correct herself. A self-repair usually comprises an erroneous item, a self-interruption, and a repair (Levelt, 1983). Sentence 1 below exemplifies a repair.

(1) Move to the red spot uh…no to the blue

Sentence 1 displays a lexical error repair, where the lexical item “red” is mistakenly articulated in lieu of “blue”. Errors are detected and repaired due to the monitoring component of speech production. Among various approaches to monitoring and to speech production in general, Levelt’s (1989; 1999) blueprint of the speaker is still one of the most widely acknowledged models of modular processing. It contains four main ingredients: conceptualization, formulation, articulation, and monitoring.

Lately, Levelt’s model has become relevant also in a bilingual perspective, such as within the field of Second Language Acquisition (SLA). Research evidence supports both similarities and differences between first language (L1) and second language (L2) users’ speech production. Some differences regard speech errors and repairs. Studies have shown that elementary L2 learners repair more frequently than advanced learners (Van Hest, 1996), and that they often employ different repair types. Advanced L2 learners make fewer lexical repairs (error repairs), and more discourse level repairs (appropriateness repairs) (Kormos, 2006). These results may relate to the amount of cognitive attention involved in speech production at various proficiency levels; lower proficiency entails more cognitively demanding processing, and consequently more errors. Levelt (1989, p. 20) refers to levels of attention with the terminology “executive control” and “automaticity”, and suggests that the former is involved in conceptualizing and monitoring while the latter is involved in formulation and articulation. He explains that while automatic functions are quick, effortless and do not compete with the other levels of processing for capacity, controlled functions are slow and they require attention, which constrains other processes. In line with this, Paradis’s (2009) neurolinguistic bilingual speech production model suggests that L2 learners initially rely on non-automatic explicit knowledge, and that automatized implicit competence accompanies increased proficiency.

The present study compares patterns of speech errors and repairs generated by L1 and L2 speakers. Note that the terminology “L2 speaker” here functions as an umbrella term for any speaker who is a successive multilingual. Levelt (1982; 1983) collected and
analysed repairs by having L1 informants describing multi-coloured interconnected node patterns. The methodology proved to be very fruitful and Levelt’s repair classification system is still applied in up-to-date studies. However, his test has only been carried out with L1 informants. Since individual multilingualism is a widely spread feature (Trudgill, 2000), it seems adequate to add a bilingual variable to Levelt’s study in order to variegate his results and to contribute to the research field of bilingual speech production. My study replicates Levelt’s methodology and presents a comparative analysis; errors and repairs were collected from a group of L2 English speakers in order to investigate the following research questions:

1. Do L2 speakers repair more frequently than L1 speakers?
2. In comparison to L1 speakers, do L2 speakers leave a larger proportion of their lexical speech errors non-repaired?
3. Do L2 speakers produce different proportions of repair types in comparison to L1 speakers?

The 3 research questions are accompanied by 3 hypotheses:

1. L2 speakers do repair more frequently than L1 speakers.
2. L2 speakers leave a larger proportion of their lexical errors non-repaired.
3. L2 speakers produce different proportions of repair types in comparison to L1 speakers; they make more error repairs and fewer appropriateness repairs.

In addition, the L2 informants’ performances are considered in relation to their proficiency levels, which are measured according to their results on a diagnostic test provided by the English Department at Stockholm University. In further addition, 6 English L1 speakers contribute as a third group of informants. Data from the two English groups is analysed and compared with data from Levelt’s group, and discussed in relation to the theoretical background.

2. Background

2.1 An L1 speech production model

Levelt’s (1989; 1999) blueprint of the speaker includes the following levels: conceptualization, formulation, articulation, and monitoring. These four machineries operate with the speaker’s individual body of conscious and unconscious knowledge, which is divided over three knowledge stores. The first level conceptualizer uses information from the store labelled “knowledge about the internal and external world” (Levelt, 1999, p. 87) in order to generate a pre-verbal message appropriate to the speaker’s communicative intention in the given discourse. This message functions as input to the second level formulator, where it is linguistically suited with information from the knowledge stores labelled “mental lexicon” and “syllabary” (Levelt, 1999, p. 87). Finally, the now phonetically encoded message reaches the third level articulator, which directs various muscles in the body to produce the spoken output. The stages of conceptualizing, formulating, and articulating are illustrated in figure 1.
Figure 1 further illustrates the monitoring stage, which is based on the perceptual loop theory (Levelt, 1983; 1989). It shows that the conceptualizer controls the monitoring processes; these processes monitor the conceptualization, and both the internal and the overt speech. Monitoring of the conceptualizer involves evaluating the pre-verbal message for appropriateness, politeness, and the order of information, etcetera. An alteration at this level may often appear overtly as a hesitation, or as a repetition; this is called a covert repair. It may also appear as an interruption of speech followed by a restart with a new formulation; this is called a D-repair. Internal speech is also monitored. Internal speech is the encoded message, which has been given lexical, syntactic, and morpho-phonological values. This message, or phonological score, is
looped back to the conceptualizer via the speech comprehension system (the same one
that is used for listening to other speakers). The conceptualizer can thus process the
message as parsed speech and compare the message to the initial communicative
intention. If the segment does not appropriately convey the intended message, the
production process is repeated and the formulation is altered. These errors are often
corrected before articulated and thus repaired covertly. However, sometimes errors or
parts of errors are overtly articulated despite detection in the internal loop. This could be
due to the incremental nature of processing; the beginning of a message may have
reached the articulator before the end of the message has left the conceptualizer. Hence,
error detection in the internal perceptual loop may generate both covert and overt
repairs. External speech is also monitored; this is articulated speech. It travels through
the conceptual loop, but since errors are detected after being articulated, their repairs are
overt.

The speech production components are either automatic or controlled by the speaker.
Levelt (1989) explains that the formulator and the articulator function to a great extent
automatically. These levels of language production, including steps such as lexical
retrieval, and grammatical and phonological encoding, are uncontrollable to the speaker
because she is not aware of them. The conceptualizer, on the other hand, is under
executive control; two stages of conscious planning generate the pre-verbal message.
Macroplanning elaborates the communicative intention by determining what
information is relevant, in what order it should be conveyed, and which speech acts are
appropriate. Microplanning determines among other things in what perspective the
information should be proposed, and how the thematic roles should be distributed; these
decisions are made in accord with the given discourse and speech context. The speaker
consciously monitors the pre-verbal message, compares it with the original
communicative intention, and alters it as necessary. The two other stages of monitoring,
the loops from internal and external speech back to the conceptualizer via the parser, are
also under executive control; the speaker consciously detects and repairs errors. In brief,
conceptualizing and monitoring generally are controlled processes, while formulation
and articulation generally are automatic processes. Controlled processing requires
attention, which is limited by the capacity of the Working Memory; the speaker can
attend only to the items held there, which amount to maximum 7 (Kormos, 2006).
Automatic processes, on the other hand, use other resources, and do not compete for
attention and Working Memory capacity. Therefore, controlled processing is slow while
automatic processing is often very fast.

2.2 L2 speech production, automaticity, and attention
As seen in the previous section, one significant difference between L1 and L2 speech
production is the amount of attention required. Note that the terminology “L2 speaker”
here refers to the successive learner, who has learnt her second language subsequently
to her first language. L1 language is, as seen in Levelt’s blueprint, to a large extent
produced automatically. L2 production, by contrast, is often controlled and thus it
requires more cognitive attention. L2 automaticity increases with proficiency and
fluency; the L2 speaker must frequently repeat a grammatical rule such as English
subject-verb agreement in order to finally automatize it. Paradis (2009) presents
language as explicit knowledge and implicit competence. Explicit knowledge is the “knowing-that” (2009, p. xi), or the content of the language; it is stored in the declarative memory source. Implicit competence is the “knowing-how” (2009, p. xi), or the set of rules that operates on the content; it is stored in the procedural memory source. Implicit competence is automatic while explicit knowledge is controlled and requires attention. Paradis suggests that the automatizing of language, such as finally producing L2 English subject-verb agreement without consciously attend to the rule, is the transformation of explicit knowledge into implicit competence. After this point, the L2 speaker can adhere to the given rule as automatically and effortless as the L1 speaker.

Levelt adds a factor to the notion of automaticity; he points out that there are also automatic features of conceptualizing, which are stored in the long-term memory. He explains that “[m]any conversational skills […] have been acquired over the course of a lifetime and are quite directly available to the speaker. They are not invented time and again through conscious processing.” (1989, p. 21). This implies that an experienced language user has quick access to complete or semi-complete pre-verbal messages, and can pay less attention to conceptualization. This observation illustrates another disadvantage for the (inexperienced) L2 speaker with regard to attentional demands. Furthermore, strings of language such as collocations, idioms, and phrases may be stored as entities in the mental lexicon. These are called formulaic knowledge; Kormos (2006, p. 45) defines these formulas as “units of language that are stored and retrieved as one single unit”. She points out that the law of contiguity, which states that frequently co-occurring items tend to form long-term associations, supports the storage of formulaic knowledge up to clause level. The usage of formulas further reduces the demands on conceptualization and attention; it further indicates that experience of a language facilitates automatic speech production. An L2 speaker, who has had less exposure to the language, will be disadvantaged and required to use more attention in comparison to an L1 speaker.

In summary, various accounts for the notion of automaticity imply a higher demand on cognitive attention in L2 production due to a higher degree of controlled processing, which declines as proficiency and fluency increase. This theoretical framework underlies the hypotheses in my study.

2.3 Levelt’s study of L1 speech errors and repairs

2.3.1 Types of errors and repairs

As mentioned before, monitoring of the conceptualizer and of the internal perceptual loop often generate covert repairs because the mistakes are prearticulatory edited. This may be overtly discernable as an interruption, or as a repetition; there is no articulated repair and thus it is hard to distinguish or analyse one. But monitoring also generates overt speech errors and repairs, which are more commonly and more easily studied. Levelt (1983; 1989) analysed recordings of 53 L1 Dutch speakers describing 53 patterns of interconnected nodes in various colours. Out of the 959 repairs collected, 236 were
covert and not relevant in the present study. Out of the 723 overt repairs, approximately 399 (55 %) were E-repairs, 290 (40 %) were A-repairs, and 10 (1.5 %) were D-repairs. These repair types are presented below. The remaining 24 (3.5 %) were R-repairs and belong to a rest category; these repairs were for various reasons not possible to label.

E-repair stands for “error repair”. These repairs involve an articulated speech error, an interruption, and a proper repair. The error may be lexical, syntactic, or phonetic. Since these errors have been articulated, they are likely to result from monitoring of the external loop. However, speech errors that are partially or wholly articulated may also result from monitoring of the internal loop.

Table 1. E-repairs (error repairs). Example sentences are from Levelt (1983, pp. 53-54).

<table>
<thead>
<tr>
<th>Name</th>
<th>Type of error</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>EL-repair</td>
<td>The speaker (to some extent) articulates an erroneous word</td>
<td>“Straight on red, or sorry, straight on black”</td>
</tr>
<tr>
<td>(Lexical error)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES-repair</td>
<td>A syntactic construction is started but not completed</td>
<td>“And black … from black to right to red”</td>
</tr>
<tr>
<td>(Syntactic error)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF-repair</td>
<td>A speech sound is misplaced</td>
<td>“A unut, unit from the yellow dot”</td>
</tr>
<tr>
<td>(Phonetic error)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D-repairs are repairs that give new, or different, information. They occur when the speaker changes her mind and starts over with new information, as in sentence 2 below.

(2) We go straight on or … We come in via red, go then straight on to green. (Levelt, 1983, p. 51)

Levelt ascribes most of these repairs to linearization problems; the speaker is unsure about what order to appropriately deliver the messages in, and wants to alter the arrangement. Linearization is a feature of macroplanning, which generates the pre-verbal message. The repair can also replace pre-verbal information that is incorrect or inappropriate. Thus, D-repairs most likely result from monitoring of the conceptualizer.

A-repair stands for “appropriateness repair”. These repairs involve adjustments of the message rather than errors. The speaker may want to further specify some articulated information, or to adjust the level of terminology as appropriate for the given discourse.
(terms such as “police man” and “cop” are examples of terminology that may be suitable in different discourses). She may also monitor for coherence with previous speech, in order to keep the message syntactically comprehensible. A-repairs are similar to D-repairs in that they adjust the pre-verbal message; they are made at a conceptual level. Since both D-repairs and A-repairs result from monitoring of the pre-verbal message, a predominant amount is likely to be repaired before leaving the conceptualizer; these are covert repairs.

Table 2. A-repairs (appropriateness repairs). Example sentences are from Levelt (1983, pp. 52-53).

<table>
<thead>
<tr>
<th>Name</th>
<th>Function</th>
<th>Example</th>
</tr>
</thead>
</table>
| **AA-repair**  
(Ambiguity) | The speaker monitors for contextual ambiguities | “We start in the middle with …, in the middle of the paper with a blue disc.” |
| **AL-repair**  
(Level of terminology) | The speaker monitors for the appropriate level of terminology in terms of accuracy | “…with a blue spot, a blue disc at the upper end” |
| **AC-repair**  
(Coherence) | The speaker monitors for coherence with the previous speech | “Go you one up, is uh… come you to yellow”  
(Original language is Dutch) |
| **ALC-repair**  
(Either level of terminology or coherence) | Either an AL-repair or an AC-repair; it is not possible to determine | A speaker who has used the terminology “disc” in all previous discourse suddenly says: “…with a blue spot, a blue disc at the upper end” |

An additional and important aspect to consider is the amount of non-repaired errors; the speaker articulates an erroneous item and either does not detect it or does not bother to correct it. These are usually difficult to label as errors, since the speaker’s intentions are never transparent. A word may appear to be misused, but the speaker could still have used it intentionally. Conversely, a seemingly appropriate word may be a speech error, but cannot be confirmed as one without the speaker repairing it. However, the methodology from Levelt’s study provides a helpful advantage here. The task to describe given patterns elicits relatively natural language that is content-controlled; the informant is speaking unconstrained while her communicated message is fairly predictable. Thus, a lexical error such as naming the wrong node colour is an obvious error even if it remains uncorrected. Levelt’s L1 speakers corrected only 46 % of their colour naming errors, while 54 % remained unrepaired. Plausibly, the majority of these
were not repaired because they were not detected; monitoring is an attention-demanding task. Since attentional resources are under more strain in L2 production, L2 speakers may be expected to correct an even smaller proportion of their colour naming errors, as stated in hypothesis number 2.

2.3.2 The structure of repair

A repair often comprises three steps: a self-interruption, an editing term and an appropriate repair. These steps are illustrated in figure 2.

Figure 2. The structure of repair (Levelt, 1983, p. 45).

The first step, self-interruption, indicates according to Levelt (1983) the point of error detection. He observed that 18 % of the overt repair interruptions took place within the reparandum (error) itself, while 51% were made immediately after completing the erroneous word. The remaining 31 % of interruptions were delayed interruptions occurring one or a few words after the error; figure 2 illustrates an interruption with a delay of 3 syllables (d=3). Levelt suggests that delayed interruption indicates delayed detection. It is also plausible that the speaker delays interruption because she wants to complete the current word or phrase, as indicated by a high frequency of interruptions at constituent boundaries. However, Levelt ascribes this trend to increased levels of attention available and thus more and quicker error detections towards the ends of constituents. He promotes his idea by showing that errors involving node colours were more often corrected in phrase-final position, while the phrase-initial errors more often remained non-repaired (and probably non-detected). If attentional resources limit the number of immediate error detections, it seems plausible that L2 speakers are more limited than L1 speakers.

The second step of the typical self-repair is an editing expression. These expressions signal to the listener that old information is being rejected and replaced, and include editing phrases like “uh”, “er”, “no”, “sorry”, “I mean”, and “rather” (Levelt, 1989, pp. 482-484). The repair may also be integrated in the editing phrase, as in sentence 3 below, where the editing terms “uh” and “rather” surround the repair word “yellow”.

8
(3) “To blue, uh, yellow rather” (example from Levelt, 1983, p. 45).

In appropriateness repairs, editing phrases are less frequent due to the absence of actual errors; the appropriateness repair merely adds or alters information, which apparently do not always require a signal.

The third and final step of the self-repair is the repair. Levelt distinguishes 3 general types of repairs: instant repair, anticipatory retracing, and fresh start. Instant repairs positions the repairing word initially, directly after the interruption and the eventual editing term, as in sentence 4.

(4) “Again left to the same blanc crossing point - white crossing point” (Levelt, 1989, p. 490).

The repair word “white” is the first word uttered and instantly repairs the error “blanc”. In other cases, the speaker incorporates one or more words preceding the error. This type of repair is called anticipatory retracing and is exemplified in sentence 5.

(5) “And left to the purple crossing point – to the red crossing point” (Levelt, 1989, p. 490).

The error word “purple” is replaced with the repair word “red”, but the repair traces back to the beginning of the prepositional phrase in which “purple” is the complement. This type of repair back-traces to boundaries of various types of constructions. The third type of repair is the fresh start, in which the speaker abandons the original utterance and starts over with a new construction. Levelt distinguishes a special type of restart, which he calls prespecification. This type of repair is characterized by a fresh start that reuses old material, as demonstrated in sentence 6.

(6) “The road begins with a – or it goes on and it begins with a green crossing” (Levelt, 1989, p. 490).

In this sentence, the segment “begins with a” is reused. Prespecifications typically occur in appropriateness repairs. In summary, 3 general types of repairs are the instant repair, the anticipatory tracing repair, and the fresh start repair.

2.4 Previous studies on L2 monitoring

A number of researchers have investigated L2 monitoring. Kormos (2006) reviews findings, which show that elementary L2 users’ and advanced L2 users’ patterns of errors and repairs may diverge both quantitatively and qualitatively. Some studies (Van Hest, 1996) show that the overall number of repairs made is higher among less proficient learners. Other studies (see Kormos, 2006) show no quantitative differences but significant qualitative differences. Elementary learners seem to more often repair lexical, grammatical, or phonetic errors while advanced learners more often repair at discourse-level; the researchers apply this to the different amounts of attention
available. As proficiency increases, language skills are automatized and the speaker may formulate error-freely without the need to monitor. Instead, her attentional resources are free for appropriateness monitoring at a conceptual level. This account is in line with Paradis’s (2009) ideas, and it explains why less proficient L2 speakers generate E-repairs and more proficient L2 speakers generate A-repairs. The informants in my study use academic English daily and must be considered advanced L2 users with a relatively large proportion of automatized knowledge. However, since L1 speech is highly automatized they can still be expected to produce more E-repairs and fewer A-repairs than L1 informants, as stated in hypothesis 3. Another interesting variable is the D-repair. Kormos (2000) found that L2 learners more frequently replace their original message with a completely new one. She suggests that the speaker does not feel linguistically competent enough to deliver the intended message and therefore changes it. This idea is returned to in section 5.

3. Methodology

3.1 Material and procedure

The data in this study is gathered with a smaller-scale version of Levelt’s (1982; 1983; Melinger & Levelt, 2004) methodology used for investigating a range of matters, including self-monitoring, but also spatial descriptions and the use of gestures. Informants were audio recorded while describing visual patterns of interconnected nodes in various colours. Figure 3 demonstrates a sample pattern. See appendices A-C for the patterns used.

![Figure 3. A node pattern.](image)

The pattern in figure 3 was used as a sample when reading the informants the instructions. They were told to for each pattern start at the arrow and describe the pattern step by step so that another (imaginary) person would be able to draw the same figure based on those instructions. In order to manage the amount of details included in the descriptions, the informants were told that the person drawing the patterns would also have seen a few samples. I explicitly explained that nobody would actually draw the pattern, since the test persons in the pilot study expressed concerns about whether their performance would match the requirements, and I assessed their concerns as
unnecessary distractions from the task. On this point the instructions deviated from the instructions given to Levelt’s informants; in his original study (1982), the participants were told that someone would actually listen to the tape and draw the patterns. The purpose of that study was to investigate how the informants described spatial relations. Levelt (1983) subsequently re-used the collected data in the study on monitoring and self-repair, in which the fabrication of a pattern-drawing person may perhaps have been less relevant. However, this deviating and possibly influencing factor should be kept in mind when viewing my results and when comparing them to Levelt’s results. In the present study 18 L2 English speakers described 3 different patterns. The informants were all English students at Stockholm University, with the native tongues Swedish, Finnish, Arabic, Greek, Punjabi, German, Spanish, or Italian. Some of the informants listed English as their L3, but as mentioned in the introduction, the term “L2 speaker” will here refer to any multilingual successive English learner. The test was conducted individually over the course of 3 weeks in soundproof study rooms in the university library. Everyone except 3 had during their studies taken the English department’s diagnostic test, which is designed to measure students’ proficiency levels; the test checks for knowledge about grammar, vocabulary, collocations, and pronunciation. Subsequently, the recordings were (non-phonetically) transcribed and analysed. The analysis considers both numbers and types of errors and repairs.

3.2 Interpreting the data

The analysis of the transcripts and the recordings comprised two steps: to detect and to appropriately categorize repairs and non-repaired colour errors. The analysis entailed several assessment ambiguities; it was often not obvious if and how to label a repair. The data in the original study (Levelt, 1983) was analysed by two judges. Their initial scoring agreed in 73 % of the cases. They discussed the residual cases and placed any uncertainties in the rest category. Levelt (1983, p. 55) refers to these uncertain repairs as “so completely confused that they defy any [other] systematic categorization”, and he does not scrutinize them as a group. When analysing the data I followed the labelling criteria outlined by Levelt; these are generally illustrated in section 2.3.1. Still, possible person-related influences on the analysis must be kept in consideration. A great difference between my analysis and Levelt’s is the collected number of repairs; I found on average 2.2 repairs per pattern description, while Levelt found only 0.25. As shown in previous studies (Van Hest, 1996), less proficient speakers can certainly be expected to more frequently repair. However, the great quantitative difference in repairs collected is also an indicator of diverging analytic criteria for category inclusion. In order to control this possible methodological discrepancy, additional data was collected from a group of 6 L1 English speakers. This group produced on average 1.06 repairs per pattern description. The amount of L1 data may be too small to generalise upon, but it can be compared with the L2 performance without any methodological inconsistencies. The results of the analyses are presented and compared with each other and with Levelt’s results in section 4.
4. Results and analysis

4.1 Numbers of repairs

There are significant differences between the numbers of repairs produced by the two groups of English speaking informants. The 18 L2 informants produced a total of 117 repairs; they averaged 6.5 repairs per informant with a standard deviation of 3.8. The 6 L1 informants made a total of 20 repairs; they averaged 3.3 repairs per informant with a standard deviation of 2.7. This shows that the L2 group made more mistakes than the L1 group, and I interpret the individual variances within both groups as relatively large. These numbers are not comparable with the equivalents in in Levelt’s study due to the issue discussed in section 3.2, and also due to the variance in the extent of the two experiments; Levelt used 53 patterns and I used 3.

Each informant’s frequency of repairs is most accurately considered in relation to her total amount of spoken words. Albeit all informants received the same recited instructions with guidance to the adequate level of details required, there was still a great variance in the amount of information that they included. Sentence 7 and 8 are non-phonetic transcripts of two L2 participants describing the first move in pattern A. The transcripts indicate hesitations but not prosodic boundaries or constituent boundaries.

(7) So it starts with an orange circle and then upwards to a green circle

(8) Uh we start with the first circle which is orange so the orange circle is first and then from this orange circle there is a... line a black line going up vertically up and you get to a green circle

These excerpts represent the variance in the number of details and words included in each participant’s performance. The total amounts of spoken L2 words excluding non-verbal fillers but including interrupted and therefore fragmental words range between 217 and 1190 words. The corresponding numbers in the L1 group range between 221 and 713 words. Thus, the frequency of repairs per 100 words is plausibly an appropriate model for comparison. The L2 group produced on average 1.48 repairs per 100 words, with a standard deviation of 0.71. The L1 group averaged 0.72 repairs per 100 words, with a standard deviation of 0.73. This clearly illustrates that the L2 informants made twice the amount of repairs compared to the L1 informants. This result confirms hypothesis number 1 and may be explained by the large proportion of non-automatic declarative knowledge utilised in L2 production; a lack of attentional resources constrains the monitoring process and thus enables speech errors. This approach, however, depends on the assumption that all repaired erroneous or inappropriate items equal all produced erroneous or inappropriate items. However, the presence of a low proportion of repair may also mirror a low proportion of error detection. Consequently, the numbers of non-corrected speech errors must also be considered.
Non-corrected linguistic defects are hard to establish as speech errors because the speaker’s intentions are normally not obvious. An inaccurate grammatical item may be intended by the speaker and inaccurate due to a gap in her linguistic competence, or due to discourse factors, such as irony. Thus, by establishing the speaker’s intentions the mistakes that are in fact speech errors can be identified. This is possible with the methodology in the present study; erroneous naming of node colours and directions can be assumed as not intended since they describe the pattern inaccurately. Sentence 9 illustrates both a colour-error repair and a direction-error repair made by an L2 informant.

(9) Directly to the left uh to the right of the ye... red circle you draw a line and you make a yellow circle.

In sentence 10, both lexical errors are repaired. Levelt (1983; 1989) compared the numbers of repaired and non-repaired colour items and found that 54% of the colour naming errors were never repaired. Based on this result, I hypothesized in the present study that the L2 informants’ proportion of non-corrected colour errors would be even larger. This hypothesis, however, was not confirmed. The L2 group generated 13 colour naming errors of which only 2 (15%) were left non-repaired. The direction errors amount to 12, of which 4 (33%) are left non-corrected. However, these 4 non-repaired items were produced consecutively by the same informant on one part of pattern B, which indicates that the informant was temporarily confused regarding the notions of “left” and “right”. The L1 group made only one direction error, which was left unrepaird. Since they made also only one repaired lexical error, their data is too small to compare with the L2 group’s data or to generalise upon. Another aspect that must be kept in consideration is the motive behind a non-repair. Hypothesis number 2 assumes that a non-repair equals non-detection. Attentional resources required for monitoring are limited in L2 production, and consequently a lower rate of error detection may be expected. However, a speech error may also be detected but intentionally left unrepaired. Again, the ambiguity lies in the opacity of the speaker’s intentions. In conclusion, hypothesis number 2 expected the L2 informants to generate a larger proportion of unrepaired colour naming errors than both the L1 informants and Levelt’s L1 informants due to the limited amount of available attentional resources in L2 production; the hypothesis was not confirmed. I will return to this matter in the general discussion.

4.2 Types of repairs

Hypothesis number 3 stated that the types of repairs would also vary between the groups; the L2 informants were expected to produce a larger proportion of E-repairs and a smaller proportion of A-repairs compared to Levelt’s Dutch L1 informants, and also compared to the English L1 informants in the present study. Figures 4-5 illustrate the distribution of repair types within the L2 group and within the English L1 group. Note that the total amounts of repairs differ substantially; the L2 group produced 117 repairs, and the L1 group produced 20 repairs.
Figure 4. The distribution of repair types over 117 English L2 repairs.

Figure 4 shows that 52 repairs (44 %) are error repairs, 32 repairs (27 %) are appropriateness repairs, 17 repairs (15 %) are in the rest category, and 16 repairs (14 %) are D-repairs, which involve fresh starts.

Figure 5. The distribution of repair types over 20 English L1 repairs.
Figure 5 shows that 14 repairs (70 %) are appropriateness repairs, 3 repairs (15 %) are in the rest category, and 2 repairs (10 %) are D-repairs, and 1 repair (5 %) is an error repair.

Figures 4-5 display a significant difference in repair types made by the two groups. As hypothesis number 3 predicted, the L2 informants produced a larger proportion of error repairs while the L1 group produced a larger proportion of appropriateness repairs. These results may be explained with the difference in knowledge types used in L1 and L2 speech production. L1 users utilise more automatic procedural knowledge, and perhaps more lexicalized ready-made multiword units. Thus, there is less room for lexical mistakes. L2 users, on the other hand, use attention-demanding declarative knowledge, and plausibly they keep fewer phrases and clauses stored. Thus, their speech production is constrained by a low level of available attention for monitoring, and by the difficulty of the task to formulate in a non-native language. In summary, the results displayed in figures 4-5 confirm hypothesis number 3.

Figure 6 illustrates the distribution of repair types within Levelt’s Dutch L1 data. His informants produced in total 723 overt repairs.

<table>
<thead>
<tr>
<th>Dutch L1</th>
<th>A-repairs</th>
<th>E-repairs</th>
<th>D-repairs</th>
<th>Rest category</th>
</tr>
</thead>
</table>

**A-repair:** appropriateness repair. The speaker monitors for ambiguity, incoherence, or inappropriate terminology, and adjusts her message accordingly.

**E-repair:** error repair. The speaker repairs a lexical, syntactic, or phonetic speech error.

**D-repair:** the speaker makes a fresh start with new information.

**Rest category:** any repair that is uncertain how to label.

Figure 6. The distribution of repair types over 723 Dutch L1 repairs.

The figure shows that 399 repairs (55 %) are error repairs, 290 repairs (40 %) are A-repairs, 24 repairs (3.5 %) are in the rest category, and 10 repairs (1.5 %) are D-repairs. These results are not in line with the results in figures 4-5. Levelt’s L1 group produced a larger proportion of error repairs compared to both groups in the present study. The two L1 groups were expected to perform in a similar manner and differently from the L2 group. Instead, there are similarities between the Dutch L1 group and the L2 group.
However, several factors must be taken into consideration when interpreting this data. First of all, Levelt’s results and mine are based on experiments with different extents; Levelt used 53 patterns and I used 3. Presumably, a task with a longer duration is more cognitively demanding and therefore elicits more mistakes. Secondly, two different languages were used in the experiment. Dutch and English languages may possess different qualities that make their users more or less prone to lexical errors. In line with this, Fox, Hayashi, and Jasperson (1996) showed that English and Japanese speakers organize their repairs differently; the researchers suggest that repairs are strongly influenced by the grammar of the language used. And finally, the person categorising repairs is not the same in all analyses; there may be inconsistencies. This issue was addressed in section 3 and is returned to in the discussion.

Another significance highlighted by figures 4-6 is the high proportions of English L1 and L2 D-repairs in comparison to the Dutch proportion. A D-repair is a fresh start that includes new, or different, information. The L2 and English L1 informants’ proportions of D-repairs amount to 15% and 10%. The corresponding number in Levelt’s results is 1.5%, which is significantly smaller. Also the rest categories are proportionally larger in the present study. These matters are discussed in section 5.

In summary, figures 4-6 display a significant and expected difference in repair types made between the English L1 speakers and the L2 speakers, but not between the Dutch L1 speakers and the L2 speakers. In addition, the proportions of D-repairs are significantly larger in the present study compared to in Levelt’s study.

### 4.3 The relation to L2 proficiency

Out of the 18 L2 informants, 15 had previously taken the diagnostic test at the English Department at Stockholm University; this test is designed to measure language proficiency. Their test scores were collected and analysed for possible correlation with their performance in the present study. The scores collected range between 90 and 150 and their average value was 125. The maximum score on the diagnostic test is 160. Thus, the L2 group may be considered as relatively proficient successive L2 English speakers.

The scores from the diagnostic test and the frequencies of repairs have a low degree of correlation (-0.14) in the Excel correlation test. The degree of correlation is low also in relation to the frequency of E-repairs (0.016), and in relation to the total numbers of spoken words (-0.009).

To summarize the results, it may be concluded that the L2 group have the highest frequency of repairs. They have a higher proportion of E-repairs compared to the English L1 group but not compared to Levelt’s Dutch L1 group. The English L1 group generated a significantly low proportion of E-repairs and a significantly high proportion of A-repairs. The L2 group left 2 (15%) of their 13 colour errors unrepairoed, while the English L1 informants made no colour errors. Finally, the present study generated an abundance of D-repairs in comparison to Levelt’s study, which may be related to factors
discussed in the next section. The overall L2 performance does not seem to correlate with the informants’ proficiency levels.

5. Discussion

5.1 Limitations of the study
The results of this project show both expected and unexpected differences between L1 and L2 repair patterns. However, these results cannot be generalised upon for several reasons. First of all, there are methodological aspects that aggravate the comparison of the data involved. As discussed in section 3, the methods of analysis in the present study and in Levelt’s study are unlikely to be identical. The great variance in frequencies of repairs collected indicates either that the English-speaking informants repaired significantly more often than the Dutch-speaking informants, or that my method of analysis was more inclusive than Levelt’s. The English L1 and L2 data is comparable since it is all analysed by the same person. However, the two groups of informants differ in numbers. The L1 group amounts to 6 informants, and the L2 group amounts to 18 informants. Thus, they too are not ideal to compare. In any future study, an L1 group and an L2 group of the same size would be preferable.

The issue with different people analysing is likely to affect not only the numbers of repairs collected, but also the proportions of repair types within each group. The hypothesis of more frequent L2 error repair was confirmed when comparing the English L1 and L2 data, but not when comparing with the Dutch data. Further, the present study generated proportionally larger rest categories, which reflects my attempts to systematically exclude ambiguous repairs. The proportions of D-repairs are also substantially larger in my data compared to in Levelt’s. This result, however, may be of some significance. Kormos (2000; 2006) suggests that L2 speakers may have an additional reason for making a fresh start; they feel insecure about their abilities to correctly formulate their initially planned pre-verbal message, and consequently they interrupt it. The abundance of L2 D-repairs supports her idea.

5.2 General discussion
Some of the results are in line with the theoretical background presented in section 2. According to Levelt (1989), L1 formulation and articulation generally are automatic processes, while conceptualisation and monitoring require attention. This is mirrored in the proportions of repair types made by the L2 informants and the English L1 informants. The L2 data comprises a predominant amount of error repairs, including lexical and phonetic repairs. These mistakes are plausibly generated in the formulator, which according to Paradis (2009) is often not under automatic control in L2 production. The English L1 data, in contrast, comprises a predominant amount of appropriateness repairs. These are made at a conceptual level (Levelt, 1989) and adjust rather than correct the message. In line with Kormos’s (2006) suggestion, it may be the automaticity of the formulation process that allows the L1 speakers to direct their available attention to appropriateness monitoring. Kormos further mentions formulaic
knowledge as an L1 advantage. These stored multiword units may also contain information about their level of terminology and what discourses they are appropriate in. Levelt (1989, p. 21) explains that “an [L1] adult’s experience with speaking is so extensive that whole messages will be available in long-term memory and thus will be retrievable”. Consequently, the L1 speaker has an advantage also at a conceptual level, when assembling the pre-verbal message. The large proportion of D-repairs made by the L2 group may display their lack of stored multiword units. Levelt (1989) ascribes D-repairs to linearization problems in the conceptualiser, but Kormos (2006) claims that an L2 speaker may make D-repairs for additional reasons, which involve insecurities about her linguistic abilities. These repairs perhaps deserve to be investigated more closely in future studies.

Other sociolinguistic factors may have influenced the informants’ performances. As mentioned above, Fox, Hayashi, and Jasperson (1996) showed that a language’s grammar might strongly influence the speaker’s organization of repair. However, the researchers further point out that “it is possible that certain ’styles’ of repair are fashionable for a given language/culture or that they work in that language/culture to create a certain construction of self (as perhaps tentative, submissive, unintelligent, etc.)”. Similarly, these language-determined factors might affect the speaker’s decision of whether to repair a detected speech error. This is a possible explanation to differences between groups of informants. In addition, Fox and colleagues (2009, p. 60) found in their cross-linguistic study of repair initiations that “there is a great deal of cross-linguistic variation with respect to favored sites of initiation but […] most of the variation can be accounted for by a few simple interactional factors”. Thus, the given discourse and speech context are also plausibly influential aspects of self-repair organization and strategy. These, however, are outside the scope of the present study.

In summary, the English L1 and L2 data may reflect that L1 speech production processes are automatized to a larger extent than L2 processes, including both formulation and conceptualisation. As mentioned however, the data is too small to be generalised upon, and methodological issues constrain the analysis of the data in relation to Levelt’s results.

6. Conclusion

This paper presents an investigation of L1 and L2 monitoring and repair during speech production. Some of the results are in line with the theoretical background, suggesting that L1 and L2 monitoring processes are similar in function, but distinct regarding the amount of attention they require. The aim to variegate the results from Levelt’s (1983) study proved difficult due to unavoidable methodological discrepancies. However, the methodology seems prosperous if all data is analysed consistently. Thus, an extended study with larger and same sized groups of L1 and L2 informants would be an interesting future project.
References


Appendix A

Pattern A
Appendix B

Pattern B
Appendix C

Pattern C
Appendix D

Information sheet

Gender:_________________________________________________________

Age:_____________________________________________________________

Score on diagnostic test:___________________________________________

My first language is____________________________________________

My second language is____________________________________________

My third language is______________________________________________

Other languages:_________________________________________________

✓ I voluntarily participate in this study.
✓ I agree to be audio recorded.
✓ I agree to let the recordings be analysed for the purpose of the study.
✓ I agree to a possible publication of the study.
✓ I understand that this study does not collect or use any information that can expose my identity.

Signature________________________________________________________