CHAPTER 2

SPEECH DISFLUENCIES

2.1 Introduction

This chapter examines the relationship between the different elements in speech disfluencies and how they have been described, categorized and studied thus far. Within the broad area of speech disfluencies, research has ranged from looking at slips of the tongue and hesitation phenomena to the structure and types of disfluencies and self-repairs. The ultimate aim of many of these studies has been to gain insight into the processes of speech production. This chapter begins by briefly outlining the process of speech production to provide a framework within which, speech disfluencies can be situated. It then explores the relationship between slips of the tongue, speech errors, hesitation, and self-repairs in the context of speech disfluencies based on research done in these areas. Finally, the categorization and analysis of speech disfluencies in the literature is examined.

2.2 The Speech Plan

Figure 2.1 (adapted from Levelt, Roelofs, & Meyer, 1999, p. 3; Postma, 2000, p. 99) shows a simplified model outlining the process of speech production from the conceptualization of a message to the articulation of the message, to illustrate the "speech plan" or "programme" or "process" referred to in the literature (Boomer & Laver, 1968; Dell, 1986; Fry, 1969; Laver, 1969; Nooteboom, 1969; Postma, Kolk & Povel, 1990). The basic flow of speech production is seen as beginning with the conceptualization of the intended message. For, example, at the conceptual stratum, a
person may want to express the word for going with someone, that is, escorting as shown in Figure 2.2, which shows the process of lexical access (Levitt, Roelofs & Meyer, 1999).

The next process is referred to as the formulation process (Fromkin, 1971; Garret, 1975). It involves lemma selection, where words appropriate to the intended meaning are selected and retrieved from the mental lexicon at a rate of, according to Levitt, Roelofs, and Meyer (1999), about two to three words per second. Simply put, the mental lexicon can be thought of as a person’s mental dictionary, holding all the information of the words a person knows in a particular language (Levitt, 1989, p.6). A lemma refers to “the lexical item’s syntactic description” (Levitt, 1998). Thus, going back to the example of the intended expression of the word escorting (see previous paragraph and Figure 2.2), the formulation process of this word presumably involves the selection of this word over any other competing word.

Since the mental lexicon contains information about a lemma’s meaning, and syntactic, morphological and phonological features (Levitt, 1989; Levitt, Roelofs, & Meyer, 1999), this means that a lemma retrieved from the mental lexicon can be syntactically, morphologically and phonologically encoded. Thus, at the lemma stratum (see figure 2.2), such information for escorting might include the fact that it is in the lexical category of a transitive verb, and that it is in this instance to be used in the present tense, progressive aspect. Other information might include the morphological structure of the intended word, such as that it requires two forms – the base word, escort, and the progressive form –ing.
Then, at the form stratum, the individual phonemes required to produce the word is retrieved and assembled, followed by the process of syllabification of the word as shown in Figure 2.2. At this point, the syllable structure (nucleus and onset) and the vowels and consonants that make up the word are made available. This will finally result in an articulatory plan or gestural score (see Figure 2.1) with the relevant speech motor movements required to produce the intended word. There may be an articulatory buffer zone, whereby speech is held while waiting to be articulated (see Figure 2.1). The movements of the relevant speech organs will finally result in the actual production of the message, which is audible to the speaker and other listeners.

Figure 2.1 shows that the processes of lexical access, syntactic and phonological encoding occur sequentially one after another. One phenomenon that appears to support such a process is the Tip-of-the-Tongue (TOT) phenomenon. TOT is a state where “a person can access semantic and syntactic properties of the TOT word and partial phonological properties such as initial sound or number of syllables, although the complete phonology remains inaccessible” (James & Burke, 2000, p. 1378). This means that speakers may be able to describe the meaning of a target word and provide syntactic properties of the word. For example, research in the area of TOT has shown that speakers are able to provide syntactic information about a target word, even though they are unable to actually produce it. Such information includes the gender of the target word in Italian (Vigliocco, Antonini, & Garrett, 1997), and whether the target word is a countable or mass noun in English (Vigliocco, Vinson, Martin & Garrett, 1999). The fact that speakers are able to do this suggests that they may have retrieved the lemma of the target word but had yet or failed to phonologically encode it.
CONCEPTUALIZER

Preverbal message

LEMMAS

LEMMATIZATION

Lemma

MORPHOLOGICAL ENCODING

Morpheme

PHONOLOGICAL ENCODING

Syllabification

Phonological word

PHONETIC ENCODING

Phonetic gestural score

ARTICULATORY BUFFER

ARTICULATION

SPEECH MOTOR MOVEMENTS

AUDIBLE SPEECH

(From Levelt, Roelofs, & Meyer, 1999, p. 3; Postma, 2000, p. 99)

Figure 2.1
Simplified Model of the Speech Plan
(Levelt, Roelofs & Meyer, 1999, p. 4).

Figure 2.2

Fragment of the Lexical Network Underlying Lexical Access from Levelt, Roelofs & Meyer (1999)
It should be emphasized that the processes explained and illustrated in Figures 2.1 and 2.2 are far more complex than thus far described. As a result, the process of speech production does not always proceed smoothly. Problems may occur at the different levels or sub-processes of the speech plan, resulting in various types of speech disfluencies. In relation to this, different theoretical models have emerged in an attempt to explain the workings of these sub-processes, as well as the occurrences of speech disfluencies, such as slips of the tongue, speech errors and hesitation.

2.3 Slips of the Tongue

In relation to speech production, slips of the tongue are seen as the result of “transient malfunctions of the sequencing system of the speech production process, which obey stringent linguistic constraints” (Boomer & Laver, 1968, p. 130). With reference to Figure 2.1 and 2.2, such malfunctions can occur at the different stages of speech production, resulting in related errors, such as phonological, lexical and grammatical errors (see Table 2.1). This is why slips of the tongue are generally considered as a form of speech disfluency, although Lickley (1994), only considers them as disfluencies if they are detected and repaired immediately.

However, as the word tongue in slips of the tongue suggests, slips are generally used to refer exclusively to phonological errors, which occur when a different sound segment is produced than from the one intended. Boomer and Laver (1968, p. 127) found that a slip could be traced to what they called “the interfering sound”, usually within the same tone-group. Such errors are similar to what Nooteboom (1969, p. 146) refers to as “errors in the programme” or the speech plan as shown in Figure 2.1. In slips, two units
of sound segments or morpheme or words interfere with each other, resulting in an error. For example, in the following utterance taken from Boomer and Laver (1968), the insertion of \( r \) in \textit{frunds} is said to be due to the interference of the word-initial consonant cluster in the word \textit{frozen}.

\begin{quote}
Both those \textit{frunds ... funds have been frozen}.
\end{quote}

(Boomer & Laver, 1968, p. 124)

This is generally referred to as an \textit{anticipatory error}, since the error is influenced by an upcoming sound, syllable, morpheme or word, even though that unit of speech has not actually been produced at the point that the error is articulated. The fact that an error is the result of an unarticulated sound suggests that this forthcoming sound and the word it is in must have already been phonetically encoded (see Figure 2.1), and was waiting for its turn to be articulated.

Such errors contrast with \textit{perseveratory errors}, where an item is affected by a previous unit, such as in the following example from Nooteboom (1969).

\begin{quote}
\textit{I prefer to preserve} (instead of \textit{I prefer to reserve})
\end{quote}

(Nooeboom, 1969, p.147)

In this example, the word-initial consonant cluster \textit{pr} in \textit{prefer} resulted in the speaker using the same cluster, making him say \textit{preserve} instead of \textit{reserve}.
2.4 Speech Errors

Slips of the tongue like the examples from Nooteboom (1969) are sometimes also referred to as speech errors (Dell, 1986), or as a category of speech errors (Boomer & Laver, 1968; Laver, 1969). In fact, as mentioned previously, speech errors are often referred to as slips of the tongue, even though the errors in themselves may not necessarily be due to errors in articulation (Fry, 1969, p. 157-158). Instead, the errors may actually occur at different stages in the process of encoding and producing speech as shown in Figure 2.1, and involve sound segments, morphemes, and words (Boomer & Laver, 1968; Dell, 1986), as presented in Table 2.1 taken from Postma (2000). This table shows the kinds of errors and repairs that speakers make. All, but the last two examples in Table 2.1, show that speakers stop their utterance soon after or immediately upon producing the error. In some cases, they hesitate by using the filled pause uh, while in others, they repair their utterance soon or immediately after interrupting their utterance.
Table 2.1

Types of Errors and Self-Repairs in Speech Production from Postma (2000)

<table>
<thead>
<tr>
<th>Example</th>
<th>Type of Error</th>
<th>Type of self-repair</th>
<th>Basic Repair Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;We start in the middle with - in the middle of the paper with a blue disc.&quot;</td>
<td>Conceptual error</td>
<td>Appropriateness repair-reformulation</td>
<td>Major retracing and revision based on syntactic structure of the utterance, leading to a grammatically well-formed continuation. Pronominalization is possible in the repair.</td>
</tr>
<tr>
<td>&quot;John comes - uh - likes to come to the party.&quot;</td>
<td>Syntactic deadlock</td>
<td>Syntactic restructuring/reformulation</td>
<td>Retracing and reformulation based on syntactic structure.</td>
</tr>
<tr>
<td>&quot;Left of purple is - uh - of white is purple.&quot;</td>
<td>Lexical error</td>
<td>Lemma substitution</td>
<td>Retrace span is morphophonologically determined. Substitute only erroneous lemma, keep the other elements.</td>
</tr>
<tr>
<td>&quot;A nun - unit from the yellow dot.&quot;</td>
<td>Phonemic error</td>
<td>Phonemic error repair</td>
<td>Minor retrace and restart.</td>
</tr>
<tr>
<td>&quot;... from my PROsodic - prosodic colleagues&quot;</td>
<td>Prosodic error</td>
<td>Suprasegmental error repair</td>
<td>Retrace and restart, mainly of lexical stress errors for pragmatic or semantic reasons</td>
</tr>
<tr>
<td>&quot;wifftidly – fully”</td>
<td>Morphemic errors</td>
<td>Repair on the fly (no clear cut-off)</td>
<td>Replace, no retrace.</td>
</tr>
<tr>
<td>&quot;...ideals by - uh - by va -voting for the Meech lake accord.”</td>
<td>Phonemic error (preceded by hesitations)</td>
<td>Repair on the fly (no clear cut-off)</td>
<td>Minor retrace and restart.</td>
</tr>
<tr>
<td>&quot;to a - uh - stapler.”</td>
<td>Unknown</td>
<td>Covert repair</td>
<td>Postponing</td>
</tr>
<tr>
<td>&quot;to - to the right.”</td>
<td>Unknown</td>
<td>Covert repair</td>
<td>Restart</td>
</tr>
</tbody>
</table>

(Postma, 2000, p. 103)

The last two examples in Table 2.1 contrast with the rest of the examples. This is because there are no overt errors in these two examples. In the first of these, the speaker resumes his utterance after hesitating with a filled pause (uh). In the second of these, the speaker repeats the word to before resuming his utterance. Thus, the type of error is 'unknown' as indicated in Table 2.1. However, since the speakers have produced some
form of disfluency, that is, a filled pause and a repeat, it is assumed that they have made a repair. This repair is deemed to have been done prearticulatorily or covertly, hence, the term ‘covert repair’ (covert repairs will be discussed in 2.8.2). In other words, the error is assumed to have been corrected prior to articulation.

2.5 Hesitation Phenomena

The relationship between slips of the tongue, speech errors and disfluencies is blurred partly because of their different definitions of terms. For example, some interpretations of disfluencies include “articulation errors” (Shriberg, 2001) or “speech errors” (Shriberg, 1994, 1996), while others see them as “slips of the tongue” (Gósy, 2001) or “mispronunciations” (Bell, Eklund & Gustafson, 2000). Types of hesitation in speech, such as silent pauses, filled pauses and prolongation of sounds, are also often considered as disfluencies. Some studies on hesitation phenomena focus solely on pauses, examining pause duration, distribution, and the grammatical or communicative functions of pauses (Goldman-Eisler, 1958a, 1958b; Griffiths, 1991; Hawkins, 1971). On the other hand, other studies look at filled pauses, realized as /æ, ɛ, ə, r, θ, m/, and repeats and false starts as part of hesitation phenomena (Beattie, 1977; Boomer, 1965; Goldman-Eisler, 1961, 1968; Maclay & Osgood, 1959).

There are also studies where hesitation in speech is studied in terms of pauses and non-silent pauses, with the latter comprising filled pauses, repeats and lengthened or prolonged syllables (Abou Haidar, 2001; Duez, 1982). Other studies examine types of hesitation like silent and filled pauses and prolongations as part of the pausing strategies that speakers use (van Donzel & Koopmans-van Beinum, 1999). Despite the many different ways of approaching and studying hesitation, the general categories of
hesitation tend to be similar. These categories are silent pauses, filled pauses, lengthenings or prolongations and even repeats.

Based on the assumption that hesitation can be regarded as a speaker’s “delaying tactics” as well as “outright devices to permit error correction”, Hieke (1981, p.148) proposed a new taxonomy for examining hesitation. His taxonomy includes the different types of hesitation, and is divided into two categories: stalls and repairs. As Figure 2.3 shows, the forms of hesitation used by a speaker to delay or stall his speech include silent pauses, filled pauses, prospective repeats and syllabic prolongations. On the other hand, retrospective repeats and false starts are said to fulfil a repair function in Hieke’s taxonomy. According to Hieke (1981), the errors being repaired could be pronunciation, semantic or syntactic errors. It is interesting to note that Hieke (1981) categorizes two types of repeats, each one fulfilling a different function (this is discussed further in 2.9.2.1.1).

![Figure 2.3](image)

Hieke's (1981) Taxonomy of Hesitation Phenomena
2.6 Disfluencies

The presence of errors, thus, seems to be associated with hesitation phenomena (Postma, Kolk & Povel, 1990). However, the way in which errors and hesitation are actually considered to be part of disfluencies, and indeed how disfluencies are actually categorized varies in the literature. Postma, Kolk and Povel (1990, p. 19) distinguish “disfluencies” from “speech errors”. Following Dell’s (1986) interpretation of speech errors, they regard disfluencies as “interruptions of the execution of a speech plan, rather than deviations from this plan” (italics added), which they consider speech errors to be.

In other words, whilst disfluencies which in this case are essentially hesitation features, such as silent and filled pauses, repeats and prolongations, are said to disrupt the flow of speech, errors such as phoneme and word transpositions, word blends, phoneme omissions are considered as instances when something has gone wrong. These problems could occur for example, at the lemma selection, grammatical or phonological encoding or articulatory stage of the speech production process shown in Figure 2.1. Based on such an interpretation, they concluded from their study that disfluencies occurred as responses to repair errors (Postma, Kolk & Povel, 1990).

2.7 The Relationship between Errors and Hesitations

The relationship between slips of the tongue (mainly articulatory errors), other speech errors and hesitations becomes clearer if the Main Interruption Rule (Nootenboom, 1980), which states that a speaker will “stop the flow of speech immediately when detecting the occasion of repair” (Levent, 1983, p. 57), is taken into account. The main point of the Main Interruption Rule is that speakers will interrupt whenever and
wherever an error is detected, regardless of phonetic, lexical or syntactic boundaries (Levent, 1989, p. 498). Errors are corrected because they are judged according to external and internal criteria (Postma, 2000, p. 101). The external criterion might, for example, be based on the linguistic rules of the language, while the internal criterion is essentially based on an individual’s perception of what constitutes an error in the particular context in which he is speaking.

In principle, when an error or a problem is detected, speakers may hesitate by lengthening or prolonging a sound or syllable or interrupting their speech by keeping silent or using filled pauses, or a combination of these (Schegloff, Jefferson & Sacks, 1977, p. 367). Thus, errors and hesitation phenomena can be seen to relate to each other in the sense that when a slip or error is detected, speech could be interrupted by one or more types of hesitation devices. This relationship between speech errors (including slips) and hesitation is shown in Figure 2.4.

![Figure 2.4: Relationship between Speech Error and Hesitation](image)

Disfluencies, therefore seem to be regarded as a superordinate category or an umbrella term (Eklund, 2003) encompassing speech errors, hesitation phenomena and even repairs. In fact, a survey of types or categories of disfluencies in the literature shows that it includes errors, hesitation devices, as well as repairs (see Table 2.2), which is why this thesis will use the term disfluencies to encompass all manifestations of errors hesitations and self-repair.
<table>
<thead>
<tr>
<th>SOURCE</th>
<th>CATEGORIES OF DISFLUENCIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adda-Decker et al 2003</td>
<td>Filler Words, Repetitions, Restarts (including Revisions)</td>
</tr>
<tr>
<td>Bell, Eklund &amp; Gustafson, 2000 &amp; Eklund, 1999</td>
<td>Filled Pauses, Unfilled Pauses, Prolongations, Explicit Editing Terms (e.g. <em>sorry, I mean</em>), Truncations (i.e. interrupted words), Mispronunciations, Repairs (substitutions, repeats, insertions, etc.)</td>
</tr>
<tr>
<td>Gósy, 2001</td>
<td>Silent Pauses, Hesitation, Repeats, Prolongations, Alterations and Changes (structural and semantic), False Starts and Slips of the Tongue</td>
</tr>
<tr>
<td>Lickley, 2001</td>
<td>Repeats, Substitutions, Insertions, Deletions</td>
</tr>
<tr>
<td>Menyhárt, 2003</td>
<td>Change of Word, Contamination, False Start, False Word Activation, Filler, Hesitation, Phonological Error, Prolongation, Repetition, Restart, Restart with Morphological Change, Serial Order Error, Silent Pause, Silent Pause within the Word, Slip of the Tongue, Syntactic Error</td>
</tr>
<tr>
<td>Postma, Kolk &amp; Povel, 1990</td>
<td>Filled Pauses, Silent Pauses, Repeats (phonemes, syllables, words), Prolongations</td>
</tr>
<tr>
<td>Postma &amp; Nordanus, 1996</td>
<td>Lexical and (Sub)Phonemic Slips, Repeats, Prolongations, Marked Blockings, Sound Interjections</td>
</tr>
<tr>
<td>Rodríguez, Torres &amp; Varona, 2001</td>
<td>Lengthenings, Silent Pauses, Filled Pauses, Lexical Disfluencies, Abandoned Sentences, Retracing (repeats, substitutions, insertions, deletions), Discourse Markers</td>
</tr>
<tr>
<td>Shriberg, 1994 &amp; 1996</td>
<td>Filled Pauses, Repeats, Substitutions, Insertions, Deletions, Speech Errors (misarticulations)</td>
</tr>
<tr>
<td>Shriberg, 1999</td>
<td>Filled Pauses, Repeats, Repairs, False Starts</td>
</tr>
<tr>
<td>Shriberg, 2001</td>
<td>Filled Pauses, Repeats, Deletions, Substitutions, Insertions, Articulation Errors (correspond to speech errors)</td>
</tr>
<tr>
<td>Stolcke &amp; Shriberg 1996</td>
<td>Filled Pauses, Repeats, Deletions</td>
</tr>
</tbody>
</table>
If disfluencies are seen as a superordinate category, then the structure of a disfluent utterance, which consists of the following parts, becomes clearer (Lickley, 1998; Shriberg, 1999):

- **Reparandum** - the part of speech which is changed or contains within it an error
- **Interruption Point** - the point when speech is halted
- **Editing Phase** - comprising silent or filled pauses
- **Repair** - the correction of the error or problem.

This structure mirrors Levelt’s (1983) structure of repairs (see 2.8.1), and the following utterances from Lickley (1998) exemplify this structure.

**Example 1**

```
just
\[\uparrow\]
Reparandum

\[\uparrow\]
Interruption
Point

\[\uparrow\]
Repair

ehm just about an inch above the starting cross

\[\uparrow\]
Filled Pause
```

**Example 2**

```
I don’t suppose you’ve got the balloons
\[\uparrow\]
Reparandum

\[\uparrow\]
Interruption
Point

\[\uparrow\]
Repair

the baboons
```

In the examples above, the reparandum contains the error or problem, while the filled pause that follows the interruption point (Example 1) is the hesitation. Hesitations like such forms of field or silent pauses are generally regarded as being part of the *editing*
phase of a disfluent utterance (Levelt, 1983; Shriberg, 2001). The repair in the examples contains the corrections of the error. As the first example shows, the repair can also comprise a repeat of a preceding word, although such a reparandum is obviously different from the second example, where an actual correction is subsequently made. Thus, as indicated in Table 2.2, the former is sometimes regarded as a type of covert repair (see 2.8.2).

In another similar definition, disfluency is referred to as instances where “a contiguous stretch of linguistic material must be deleted to arrive at the sequence the speaker ‘intended’...” (Shriberg, 1994, p.1). This is illustrated in the following utterance, where the four disfluent portions are crossed out:

\[\text{they} \quad \text{they basically reviewed oregon’s-plan \_\_ \_ \_ \_ the oregon plan toward \_\_\_ nationalizing health care} \]

(Example from Shriberg, 1994: 1)

The parts that are deleted correspond to Lickley’s reparanda, except for the filled pause \textit{uh}. The parts that are deleted in Shriberg’s example are the first part of a repeat (\textit{they they}), and the items, which in the repair are preceded by an insertion and then repeated (\textit{oregon’s-plan ore the oregon plan}). As in the previous examples, the interruption point in the example above precedes the Editing Phase (Shriberg, 1994; 1996).

Following such a structure of a disfluent utterance, the relationship between speech errors and hesitation illustrated in Figure 2.4 can be expanded to include repairs as shown in Figure 2.5 below.
Figure 2.5

The Relationship between Speech Error, Hesitation and Self-Repair in a Disfluent Utterance

Figure 2.5 shows that the types of hesitations found in the editing phase need not always be present in disfluent utterances (Blackmer and Mitton, 1991). This is why Lickley (1994) prefers not to use the generic term “hesitation” or “pause”, to refer to disfluent utterances as these suggest that there must necessarily be a temporal break, such as a silent pauses or indeed any other form of hesitation in a disfluent utterance. Instead the blanket term “disfluency” is preferred as this includes all instances of disfluent utterances which may or may not have a filled or silent pause in the editing phase.

2.8 Self-Repairs

As shown in Figure 2.5, a repair can be considered as a part of the structure of disfluencies, and indeed, Heeman and Allen (1994, p. 295) consider speech repairs as disfluencies. This is logical given that they are produced in response to errors as can be seen in Table 2.1 (taken from Postma, 2000, p. 103), which shows some of the types of errors and their corresponding repairs. In fact, self-repairs appear to be a normal phenomenon in spontaneous speech, where speakers regularly correct themselves while speaking. For example, Nooteboom (1980, p. 94) found that 64% of speech errors in his corpus were corrected by the speakers.
Repairs associated with disfluencies are often referred to as *self-repairs*. Self-repairs, within the context of disfluencies, occur within the same speaking turn, and are "self-initiated" (Schegloff, Jefferson & Sacks, 1977, p. 366). Nakatani and Hirschberg (1994, p. 1607) consider repairs to be "the self-correction of one or more phonemes (up to and including sequences of words) in an utterance". A self-repair is generally examined in relation to a preceding error (Levelt, 1983; Postma, 2000) to determine the type of error and repair being made (see Table 2.2). An example of a self-repair from the data in this study as follows:

C29
\[ \text{C when [they] want IP when you want(SUB)} \]

The self-repair in this example shows that the speaker is substituting the word *they* with *you*. The repair involves the repetition of the words before and after the substituted word. In other words, repairs can involve a retrace to an earlier point in speech. This is why van Wijk and Kempen (1987), who, similar to Maclay and Osgood (1959), looked at repairs in terms of the relationship between the repair and the original utterance, divided repairs into "retracing" and "non-retracing" repairs. Retracing repairs, such as in the example given, occur when speakers interrupt their speech, and go back to an earlier point in their speech, repeating their utterance fully or partly. On the other hand, in non-retracing repairs, the repair replaces the reparandum without any backtracking. The following utterances from van Wijk and Kempen (1987, pp. 405 & 431) exemplify these two types of repairs.

(i) \[ \text{We start [right] on the somewhat right of the paper} \]
   \[ \text{(retracing repair)} \]
   \[ \text{(n.b. the square brackets indicate the point of retracing)} \]
(ii) *I want you to be happy ... tonight.* (non-retracing repair)

Other instances of retracings would be repeats such as in the following example:

C8
C it's [quite] IP m quite(REP) correct.

### 2.8.1 Structure of Self-Repairs

Although it has been shown that utterances that contain disfluencies can be examined in different ways (Lickley, 1994), the structure that is adopted and adapted in most studies on self-repair in disfluent speech is the structure put forward by Levelt (1983). The components of the structure are essentially the same, even if the terms used to describe it differ, such as the *structure of repairs* (Blackmer & Mitton, 1991; Levelt, 1983; Nakatani & Hirschberg, 1994) and the *structure of disfluencies* (Lickley, 1998; Shriberg, 1999).

Levelt (1983, p. 45) found that most repairs comprised three parts: the original utterance, the editing phase and the repair. This structure of repair is shown in Figure 2.6. The first part is the "original utterance (OU)", which has within it the reparandum or the problem as shown in Figure 2.6. According to Levelt (1983), the reparandum can range from a single sound to a unit of text. Similarly, Lickley (1994) and Shriberg (1999) regard the entire section of speech that the repair replaces as the reparandum.
The Structure of a Repair (Levelt, 1983, p. 45)

Levelt's original utterance begins from "the last sentence boundary before the reparandum to the moment of interruption (I)" (1983) as can be seen in Figure 2.6. Shriberg (1994, p. 6) refers to the moment of interruption, that is, the point where speech is halted, as the "interruption point (IP)", while Blackmer and Mitton (1991, p. 180) call this interruption the "cut-off". For the purposes of this thesis, both the terms interruption point or IP and cut-off will be used.

In the example above, the word left is regarded as the error since it is replaced by the word pink in the repair. The moment of interruption precedes the editing term, uh, and can thus be determined to begin upon the offset of the last segment in the original utterance phase. Even if there is no editing phase, there are said to be particular acoustic cues that signal the offset of the reparandum. These cues are as following (Nakatani & Hirschberg, 1994):
• word fragmentation
• glottalization
• coarticulation

Nakatani and Hirschberg (1994) found that 73.3% of all the reparanda in their corpus ended in word fragments, such as in flying to Sa-. They found that a large majority of these fragments (91%) were one syllable or less in duration. Other studies have also found similar evidence (Bear, Dowding & Shriberg, 1992; Blackmer & Mitton, 1991; Nooteboom, 1980; Shriberg, 1994). If present, word fragmentation can be considered a reliable cue for the end of the reparandum since it is at this point that a word is being cut-off, and therefore, speech interrupted.

Glottalization was also found in 30.2% of reparanda offsets in Nakatani and Hirschbergs’ (1994) corpus, suggesting that this is another acoustic cue that can be used to mark the end of the reparandum (Nakatani & Hirschberg, 1994, p. 1608). Nakatani and Hirschberg also found the presence of coarticulatory gestures at the offset of the reparanda in their study. Similar findings were reported in Bear, Dowding & Shriberg (1992) and Shriberg (1999). Such coarticulatory cues, according to Nakatani & Hirschberg (1993) can help to determine the word the speaker intended to produce before the moment of interruption. This phenomenon is present because the articulators would already be in motion to produce the next sound segment, and thus, even if that sound is not produced, some of its articulatory features, such as lip rounding, may be discernable. Thus, the word the speaker cut off may be determined based on such coarticulatory clues as well as the context in which the cut off word is produced.
n interruption point can be followed by one or more of the following (Lickley, 1998; Shriberg, 1999, p. 619):

- a silent pause
- a filled pause
- a lexical editing term (for example, *I mean*)

These correspond to the second part of the repair in Levelt's (1983, p. 45) structure of repair, the "editing phase" (see Figure 2.6), which need not necessarily have within it an "editing term (ET)". Thus the editing phase, which begins from the interruption point to the beginning of the repair, may be empty. For example, Nakatani and Hirschberg (1994) found that only 9.4% of the self-repairs they were examining contained filled pauses. It is interesting to note that Blackmer and Minton (1991, p. 180), include editing terms, which they defined as "filler[s] without semantic value in relation to the sentence under construction, and without function as a discourse marker..." as part of the repair.

The repair (R) is the third part in Levelt's repair structure (see Figure 2.6). The repair generally contains within it the correction of the reparandum. Heeman, Loken-Kim and Allen (1996, p. 362) and Heeman and Allen (1999, p. 528) refer to this section as the "alteration". Shriberg (1994, p. 9) refers to it as the "repair region", which seems more appropriate given that the repair, if there is one, does not make up the entire segment, but is situated within this part of the utterance. Both Lickley (1998, p. 21) and Shriberg (1999, p. 619) separate this region into the actual "repair" and "continuation". The latter is the part of the utterance, which follows the actual repair.
It should be noted that there may not actually be an overt reparanda, and subsequently no corresponding repairs, even though there may be filled pauses in the editing phase, such as in the following examples from Shriberg (1999, p. 619).

<table>
<thead>
<tr>
<th>Prior Context</th>
<th>Reparandum</th>
<th>Editing Phase</th>
<th>Repair</th>
<th>(Continuation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>it's</td>
<td></td>
<td>um</td>
<td></td>
<td>we’re fine</td>
</tr>
<tr>
<td></td>
<td></td>
<td>uh</td>
<td></td>
<td>after five</td>
</tr>
</tbody>
</table>

In other words, disfluencies can comprise only the editing phase (Shriberg, 1999, p. 619), which is presumably why this phase is not included as part of the repair in most studies (compared to Blackmer & Mitton, 1991). Instead it is seen as the portion of speech that follows the interruption point and precedes the repair. Shriberg (1999) also points out that the use of the word “editing” does not imply that errors are detected since there are other reasons for silent and filled pauses to occur. Levelt (1983, p. 44) refers to this as this as “zero alteration”, where the speaker continues his utterance after the interruption point and the editing phase. This is said to be an instance of a covert repair (Levelt, 1983).

2.8.2 Covert Repairs

Levelt (1983) goes on to say that if there are no alterations, that is, “no morphemes are changed, added or deleted”, then this is a “covert repair” (1983, p. 45), and they are said to be characterized by an interruption point (marked by, ) plus ET (example 1), or a repeat (example 2) as shown in the following Dutch examples from Levelt (1983, p.55).

Example 1

Dan recht, saf, uh grijs

*The right, uh grey*
Example 2

En aan de rechterkant een oranje stip, oranje stip
And at the right side an orange dot, orange dot

Blackmer and Mitton (1991, p. 182) expanded on Levelt's definition of covert repairs, and divided within-utterance covert repairs into three sub-categories to mirror what they saw as being different speech processes, although what these processes are were not adequately explained. These sub-categories include repeats, both with and without editing terms, and utterances with only within-utterance editing terms and no overt errors or repeats, as illustrated in Table 2.3. They also had a separate category for between-utterance covert repairs.

Table 2.3

Sub-Categories of Covert-Repairs

<table>
<thead>
<tr>
<th>Within-utterance covert repair</th>
<th>Within-utterance editing term e.g. most of us/uh do evolve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetition without editing term e.g. It was a combination of a career move.</td>
<td></td>
</tr>
<tr>
<td>Repetition with editing term e.g. and/uh and I've made a career move</td>
<td></td>
</tr>
<tr>
<td>Between-utterance covert repairs</td>
<td>Between-utterance editing term e.g. It's costing me a bundle./um I really am (I = cut-off)</td>
</tr>
</tbody>
</table>

(Blackmer & Mitton, 1991, p. 182)

Thus, it appears to be accepted that repairs can comprise only hesitation, such as prolongation, filled and silent pauses (Shriberg, 1999). Heeman and Allen (1994, p. 295) refer to such instances as “abridged repairs” in contrast with modification repairs, where there are actual changes made in the repair, and fresh starts. Covert repairs suggest that
errors are detected and corrected prior to articulation (Hockett, 1967), that is, they are prearticulatory repairs. This contrasts with postarticulatory repairs or "overt repairs", where the error or part of it and its corresponding repair is produced and can be heard. In contrast, the error and repair are not overtly observable (Postma, 2000) in covert-repairs, and as Levelt (1983, p. 55) points out, in such cases, "it is almost always impossible to determine what the speaker is monitoring for". Similarly, Hartsuiker and Kolk (2001, p. 116) stress that it is difficult to ascertain if anything is being repaired at all. Because of this, some studies omit covert repairs as a category of self-repair (Bear, Dowding, Shriberg & Price, 1993; Nakatani & Hirschberg, 1994).

Because there is no overt error and corresponding repair, repeats are sometimes put under the category of covert repairs (Blackmer & Mitton, 1991; Levelt, 1983). However, other studies (Lickley 1998; Shriberg, 1994) regard the first part or the first token in a repeat as the reparandum and the second one as the repair, and thus consider it as a form of self-repair. Others have considered repetitions as a type of hesitation together with filled pauses and prolongations rather than a type of repair, (Abou Haidar, 2001; Duez, 1982). In terms of van Wijk and Kempen’s (1987, p. 405) categories, repeats would be an example of "retracing repairs" since the speaker goes back to an earlier sound or word. (Repeats will be explored in more detail in 2.9.2).

2.8.3 Classes of Repairs

Covert self-repairs are one of the many types of self-repairs speakers make. Apart from covert repairs, Levelt (1983, pp. 51-55) distinguishes between three main different
classes of self-repairs based on the reasons speakers make them. These classes of repairs with Dutch examples from Levelt (1983, pp. 51-54) are as follows:

(i) **D-repairs** or different-repairs are made when a speaker decides that a different idea needs to be said, and so interrupts his speech to say it, for example:

We gaan rechtdoor ofte ... We komen binnen via rood, gaan dan

*We go straight on or ... We come in via red, go then*

rechtdoor naar groen

*straight on to green.*

(ii) **A-repairs** or appropriateness-repairs occur when a speaker needs to express his intended idea or message more appropriately. Therefore, such repairs are not correction of errors *per se*, but repairs to make an utterance more appropriate by correcting “potential ambiguity of reference” (AA-repairs), using more “appropriate level terminology” (AL-repairs), or to establish coherence with a previous utterance (AC-repairs). An example of an AL-repair is as follows, where the speaker substitutes *vlakje (spot)* with a more specific term, *rondje (disc)*:

...met een blauw vlakje, een blauw rondje aan de bovenkant

*with a blue spot, a blue disc at the upper end*

(iii) **E-repairs** or error-repairs are made when a speaker has detected a lexical, syntactic or phonetic error in his speech. As a result, the repairs can be lexical repairs (EL-repairs), syntactic repairs (ES-repairs) and phonetic repairs (EF-repairs). Examples of E-repairs are:
EL-Repair

Ga dan naar de verkeer ... naar de andere kant

Go then to the wrong ..., to the other side

ES-Repair

En zwart ... van zwart naar rechts naar rood

And black ..., from black to right to red

EF-Repair

Een eenheed, eenheid vanuit de gele stip

A unit, unit from the yellow dot

Levlt (1983) also had another category, R-repairs, but this consisted of a small set of repairs which were impossible to categorize.

Blackmer and Mitton (1991) adopted Levelt's (1983) classification of repairs in their study on the timing of repairs, using the two main categories of repairs (covert and overt repairs). Appropriateness-repairs and different-repairs were subsumed as subcategories of conceptually based overt repairs, while error-repairs were termed as production-based repairs (Blackmer & Mitton, 1991, p.181). These categories of repairs are shown in Figure 2.7.
Covert repair
(see Table 2.3)

Overt Repair (with or without editing terms or repetition)

Conceptually based overt repair

Appropriateness repair

Appropriateness insert
Ex: "such as Mr./uh I'm sorry to say, Mr. Broadbent's"

Appropriateness replacement
Ex: "because most/ a lot of people are"

Different repair
Ex: "I don't/ uh Bourassa is..."

Production-based overt repair
Ex: "if you some/ love something"
Ex: "you cannot even get a job in an English hospital without passing an English/ a French test."
Ex: "if Quebec can have a ba/ a Bill 101"
Ex: "behind her/ behind her own closed doors"

Unclassifiable overt repair
Ex: "who is continuing to practice after huh/ ordination"
Ex: "should be persecuted for expressing the/ that activity"

n.b. slash (/) = cut-off


Figure 2.7

Covert and Overt Repairs

2.9 Categories of Disfluencies

As mentioned earlier in this chapter, there does not appear to be a consensus as to the actual categorization of types of disfluencies even if it is generally agreed that the structure of disfluencies comprises the reparandum, the interruption point and the repair. As shown in Table 2.2, the categories of disfluencies tend to differ depending on
the focus of the study. Because of this, phenomena such as silent and filled pauses may be excluded as a type of disfluency, but included as part of the editing phase following the interruption point and preceding the repair (Lickley, 1994). In other studies, filled pauses may be categorized as disfluencies, and at the same time be considered part of the editing phase (Shriberg, 1996; 1999). The following sections will consider the different types of disfluencies that have been previously studied. The disfluencies will be looked at from two aspects: hesitation and self-repair, although it should be noted that disfluencies like repeats have been considered as instances of both these categories.

2.9.1 Types of Hesitation

The types of hesitation that have been studied are silent pauses, filled pauses and prolongations. Hesitation refers to instances when speech is interrupted by silence (for example the 227msec pause in C5), filled pauses (like er in 5) or by the prolongation of sounds within a word (like in C65) as shown in the following examples taken from the data:

C65
C ... I mean it's like: :227 something else ...

C5
C ...I like er tall guys.

Allwood, Nivre & Ahlsén (1990, p. 10) highlight the fact that disfluencies such as these three types of hesitation take up production time. This time may be indicative of planning or monitoring processes within speech production.
2.9.1.1 Silent Pauses

While it is generally agreed that a silent pause refers to "the perception of silent portion in the speech signal" (Zellner, 1994, p. 44), there are still problems in relation to the following:

- distinguishing hesitation pauses, that is, those "that are an interruption in the fluent stream of speech" (Deese, 1980, p. 71) and those which are not (e.g. grammatical pauses, stop closures or "intra-segmental pauses" or "inter-lexical pauses" (Hieke, Kowal & O'Connell, 1983; Zellner, 1994, p. 42).
- determining the cut off-point for hesitation pauses.
- deciding the threshold level where speakers have stopped speaking altogether.

There are two main issues to resolve when examining silent pauses as a means of hesitation. Firstly, hesitation pauses must be distinguished from phonetic, grammatical and rhetorical pauses (Deese, 1980, p. 71). Goldman-Eisler (1968), for example, used the following criteria to determine "non-grammatical breaks" or hesitation pauses:

1. Where a gap occurs in the middle or at the end of a phrase...
2. Where a gap occurs between words and phrases repeated...
3. Where a gap occurs in the middle of a verbal compound ...
4. Where the structure of a sentence was disrupted by a reconsideration or a false start...

(Goldman-Eisler, 1968, p. 13).
On the other hand, grammatical pauses are expected at the following points:

(1) Natural punctuation points
(2) Immediately preceding a conjunction
(3) Before relative and interrogative pronouns
(4) When a question is indirect or implied
(5) Before all adverbial clauses of time, manner and place
(6) When complete parenthetical references are made

(Goldman-Eisler, 1968, p. 13).

Once a distinction has been made between grammatical and non-grammatical pauses, the second issue has to be resolved: distinguishing between hesitation pauses and articulatory gaps, for example, caused by the closure phase of stops.

2.9.1.1.1 Measurements of Silent Pauses

The problem in terms of measurements is that there seems to be no consensus as to the exact duration of hesitation pauses. The point at which pauses are deemed to be hesitation rather than articulatory pauses range from 100msec to 250msec in the literature (Hieke, Kowal & O’Connell, 1983; Rochester, 1973). Lounsbury (1954, cited in Deese, 1980, p. 71) proposed that hesitation pauses should be more than 100msec. Similarly, Griffiths (1990, cited in Griffiths, 1991, p. 346) and Arim, Costa and Freitas (2003) used a cut off point of 100msec. On the other hand, van Donzel and Koopmans-van Beinum (1996) used a duration of at least 150msec to measure silent pauses in their study on the basis that this would exclude stop closures.
Longer cut off points have been used in other studies to define silent pauses. Boomer (1965, p. 156), for example, makes a case for a 200msec cut off point, based on the fact that listeners seemed able to distinguish a hesitation pause of 200msec. This is similar to Beattie (1977, p. 283) who used a cut off point of 200msec or more to define silent pauses. Brotherton (1979, p. 184) used a 0.250second (250msec) criterion, similar to Goldman-Eisler (1968, p. 12). The rationale for this was that any measurement below 250msec would be a phonetic stop. Duez (1982, p. 13), however, used a speaker-based criterion to measure silent pauses. Thus, depending on the speaker, Duez had thresholds of 180 to 250msec.

In contrast, instead of setting a cut off point to measure silent pauses, Lickley (1994, p. 211) measured pauses acoustically from the onset to the offset of the absence of an acoustic signal (signalling the beginning and end of the pause), allowing a closure phase of 50msec should a stop follow the pause. Doing this would mean that even pauses shorter than 100msec can be regarded as a silent pause, provided that it is not a phonetic pause. As a result, the pauses in Lickley’s study ranged from 34 to 1134msec. This seems to be a better method to use as it does not exclude shorter non-phonetic pauses, which may be theoretically relevant. This is especially pertinent as shorter non-phonetic pauses have been found in other studies (Hieke, Kowal & O’Connell, 1983; Kirsner, Dunn, Hird, Parkin & Clark, 2002; Kirsner, Dunn & Hird, 2003).

Despite the different measurements, Zellner (1994, p. 44) notes that hesitation pauses are generally taken to be in the range of 200-250msec. A speaker is generally deemed to have stopped speaking after about 3 seconds (Griffiths, 1990 in Griffiths 1991, p. 346). For instance, Goldman-Eisler (1968, p. 15) in her study, found that the pauses made by
her subjects were always less than 3 seconds, with 99% of the pauses being less than 2 seconds. Similarly the mean duration of silent pauses in Grosjean and Deschamps (1975, cited in Zellner, 1994) was below 1500msec, while the longest pause duration found in Arim, Costa and Freitas (2003) and Lickley (1994) was 1134msec and 1533msec respectively.

2.9.1.1.2 Location of Silent Pauses

Studies on hesitation pauses looked at where pauses occurred, and what this reflected. Pauses are mostly found at syntactic and prosodic boundaries (Boomer 1965; Brotherton, 1979; Hawkins, 1971), and largely because of this, they are generally associated with cognitive processes, where pauses are regarded as a manifestation of the thinking process, that is, “... hesitation pauses in speech are the delays due to processes taking place in the brain whenever speech ceases to be the automatic vocalization of learned sequences...” (Goldman-Eisler, 1968, p. 58.).

The notion that pauses are manifestation of ongoing cognitive processes can be related to the location of these pauses. For example, Maclay and Osgood (1959, pp. 27 & 39) found that pauses (filled and unfilled) occurred more frequently before “lexical words” (corresponding to nouns, verbs, adjectives and adverbs) than “function words” (articles, verb auxiliaries, connectives, prepositions etc.). This is said to be because the former is more “difficult” or cognitively demanding. This implies that the process of lexical selection for lexical words is somehow more demanding than function words. Furthermore, it is anticipated that fluent speakers of a language would not have
difficulty in the correct selection of function words for example in the correct use of articles or prepositions.

Another variable that had an effect on the rate of pauses was task difficulty. In Goldman-Eisler’s (1961) study, for example, subjects tended to have more silent pauses when asked to interpret, compared to when they were describing the cartoons they were given. The reason for this was said to be the higher cognitive task demanded by the act of interpreting, where subjects would have had to conceptualize, formulate and produce ideas and opinions of their own.

However, both Goldman-Eisler (1961) and Duez (1982) found that the ratio of filled pauses and unfilled pauses depended on individual speakers, and how they were inclined to use these pauses. Generally though, Goldman-Eisler (1961) theorized that silent pauses signalled thinking or cognitive activity, while filled pauses mirrored the emotional attitudes of the speaker, although the extent to which the above holds true has yet to be proven conclusively:

The two hesitation phenomena of filled and unfilled pauses ... appear to reflect different internal processes, cognitive activity being accompanied by an arrest of external activity (speech or non-linguistic vocal action) for periods proportionate to the difficulty of the cognitive task, while emotional attitudes would be reflected in vocal activity of instantaneous or explosive nature.

(Goldman-Eisler, 1961, p. 25)

In relation to their location in the structure of disfluencies (see 2.7), Howell and Young (1991, pp. 740 & 757) found that pauses tend to occur prior to a repair. This suggests that speakers use pauses to indicate the start of a repair. Similarly, Nakatani and Hirschberg (1994) found that the duration of pauses was a reliable marker of what they
termed the *Disfluency Interval*, that is the duration between the interruption point to the onset of repair (similar to the Editing Phase). They found that silent pauses were also significantly shorter for fragment repairs (see 2.9), where speech was cut off mid-segment, than for non-fragment repairs as shown in Table 2.4. Nakatani and Hirschberg (1994) suggest that this could indicate that speakers need less time to start the repair when they stop themselves mid-segment, which suggests that the planning of the error had been done prior to interruption. Further evidence of such prearticulatory repair-planning is the low frequency of filled pauses occurring after fragments.

**Table 2.4**

<table>
<thead>
<tr>
<th>Pause Juncture</th>
<th>Mean</th>
<th>s.d.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment</td>
<td>289ms</td>
<td>377ms</td>
<td>264</td>
</tr>
<tr>
<td>Non-fragment</td>
<td>481ms</td>
<td>517ms</td>
<td>82</td>
</tr>
</tbody>
</table>

(from Nakatani & Hirschberg, 1994, p. 1609)

Thus, pauses are not only a reflection of on-going internal cognitive and speech processing, but also serve as linguistic cues to listeners. Both silent and filled pauses are seen as strategies used by speakers to hold on to their speaking turn whilst giving them time to think of what to say or how to continue (Maclay & Osgood, 1959). In short, speakers use such pauses to stop disfluent speech and to subsequently correct them. However, it should be noted that not all repairs are preceded by a pause. Lickley (1994, p. 211), for instance, found that not all 30 utterances in his corpus had a pause between the end of the reparandum and the onset of the repair, where only 16 out of 30 disfluent
utterances had pauses. Similarly, pauses can occur as a form of hesitation without the presence of overt errors (see covert repairs in 2.8.2).

2.9.1.2 Filled Pauses

Filled pauses are generally understood to mean unlexicalized vocalizations which are commonly used by speakers as fillers in their utterances. It is interesting to note that speakers of a particular language use similar sounds as filled pauses in that language. This may be an indication that these sounds are acquired as part of the language. Examples of filled pauses in different languages are as *uh*, *ah* and *um* in English (Goldman-Eisler, 1961, p. 18); *uh* [ɔh], *um* [ɔm] and *m* [ʌm] in Dutch (Swerts, Wichmann & Beun, 1996, p. 1033); *eu, eh, oe, n* in French (Duez, 1982, p. 14); *ano* and *eto* in Japanese (Watanabe, 2001, p. 89) and *eh* and *ôh* in Swedish (Bell, Eklund & Gustafson, 2000, p. 627).

Shriberg (1999, p. 621) points out that although the vowels in filled pauses, for example, /ə/ in English, may be mistaken for a word with a similar vowel sound, such as the article *a*, the vowels in filled pauses have considerably longer durations compared to the same vowels in fluent contexts. The vowel duration of filled pauses can be acoustically measured from its waveform. Duez (2001, p. 42) for instance measured it “as the duration of the waveform between the first and the last [vowel] period”. Thus, acoustic measurements as well as the context in which they occur can be used to differentiate between a filled pause and a similar-sounding word.
Filled pauses have been shown to play a role in the making and understanding of meaning as made evident by studies that show that different filled pauses may be related to different functions. Filled pauses have been divided into short or long (lengthened) sounds (Rose, 1998). The former were also categorized according to whether they were open (uh and er) or closed sounds (um and erm). Some studies have reported a difference in the use of uh and um in relation to signalling a speech problem. For example, Fox Tree (2001, p. 325) found that a short delay was found to be signalled by uh, whereas a longer delay was signalled by um. This suggests that um tends to signal a major problem, while uh signals a minor problem. This is similar to Smith and Clark (1993, p. 35), who found that in the context of answering questions, um was followed by a pause 62% of the time, while uh was only followed by a pause 23% of the time. Thus, while a speaker more frequently resumed his speech immediately after the production of uh, while continuing to hesitate with a pause after um in most cases. This use of the two filled pauses is illustrated in the following example from Smith and Clark (1993, p. 29):

Q     In which sport is the Stanley Cup awarded?
A     (1.4) um (1.0) hockey

(n.b. the numbers in parenthesis indicate the duration of the silent pause in seconds)

Apart from the unlexicalized filled pauses mentioned thus far, discourse fillers (Fox Tree & Shrock, 1999), such as you know, well, okay and like, are often included as a category of filled pauses since not, just like unlexicalized filled pauses, such expressions can be considered as instances when a speaker hesitates while speaking. Rose (1998), for example divides filled pauses into those which are “unlexicalized”, such as er and erm, and those which are “lexicalized”, such as you know, well, okay and like. Similarly,
Abou Haidar (2001, p. 33) also includes words like *well* as categories of filled pauses, calling them “hesitation”. Lexicalized filled pauses appear to constitute a break in the flow of the utterance, and therefore can be seen as a result of the process of speech-monitoring and error-detection, where the use of such pauses indicate that speakers may have run into problems at some stage of the production process of their utterance.

The use of such lexicalized fillers as possible hesitation devices can be seen in Enfield’s (2003) study on the use and meaning of expressions like *what-d’youcall-it*. Enfield (2003, p. 104), explains that one of the interpretations of this expression is “I can’t say the word for this thing right now”. For example, in an utterance like the following one from Enfield (2003, p.104), the speaker is at first unable to produce the word he wants, resulting in the use of *what-d’youcall-it* and *you know*, after which he successfully produces the correct lexical item, *chuck key*.

*Where’s the what-doyoucall-it?..., you know, the chuck key?*

Lexicalized filled pauses, like their unlexicalized counterparts, therefore, signal that the speaker is encountering a problem. However, because they are lexicalized, they are sometimes categorized as a separate entity, and not as a filled pause. This has further implications as to their position in the structure of repairs. Heeman and Allen (1994, p. 297), for example, subsume both filled pauses and “clue words” as types of “editing terms”. Lickley (1998), however, specifically refers to phrases like *I mean* as “lexical editing terms” in contrast to sounds like *um, uh, eh, ehm* which are categorized as filled pauses. This is similar to Clark and Wasow (1998, p. 201), who refer to filled pauses like *uh* and *um* as “fillers” and expressions like *I mean, you know* as “editing expressions”. Bell, Eklund & Gustafson (2000) also refer to phrases like *I’m sorry, No*,
wrong, *I mean* as “explicit editing terms”, while Bear, Dowding, Shriberg & Price (1993, p. 5) label them as “cue words or phrases” if they occur immediately before the repair site as they are seen as a signal that there is a problem and act as a cue to the upcoming repair. Since such expressions are regarded as cues, then they could be considered as part of the repair as Blackmer and Mitton (1991) did (see 2.8.1). However, since they are generally not a part of the meaning and content flow of the utterance, they are often considered part of the editing phase of a repair.

2.9.1.2.1 Location of Filled Pauses

Maclay and Osgood (1959, p. 39) found that filled pauses occurred more frequently at phrase boundaries than within phrases, and similar to silent pauses, filled pauses appeared more frequently before lexical words than function words in their study. Cook (1971, p. 138), however, found that both filled pauses occurred equally before lexical and function words. Nevertheless, he did find that similar to Boomer (1965), filled pauses occurred more frequently after the first word and before the second word in a clause. The reason for this was said to be as follows, where filled pauses are used as a device to buy time while the process of re-planning the upcoming utterance (which must flow from the word preceding the filled pause) can take place or continue:

> The initial word in a phonemic clause sets certain constraints for the structure of what is to follow. The selection of a first word has in greater or lesser degree committed the speaker to a particular construction or at least a set of alternative constructions, and has also foreclosed the possibility of other constructions.  
> (Boomer, 1965, p. 156)

Cook, Smith and Lalljee (1974, p. 15) also found that filled pauses tended to be located either at or close to the beginning of a clause, again suggesting that this is because of the “syntactic uncertainly” at this point. Thus, filled pauses are often considered as planning
devices for upcoming utterances (Goldman-Eisler, 1972; Levelt 1989; Maclay & Osgood, 1959). In the context of speech disfluencies, filled pauses are also regarded as warning devices, signalling potential speech problems or errors (Clark 1994; Clark & Wasow, 1998; Fox Tree, 2001; Fox Tree & Schrock, 1999).

Apart from these functions, filled pauses also function as a time-buying and turn-keeping device (Maclay & Osgood 1959; Schegloff, 1981), which speakers can use to avoid being interrupted (Beattie, 1977). This is because, unlike silent pauses, filled pauses involve the vocalization of sound, and thus, indicate that the audibly indicate that the speaker intends to continue his utterance. Filled pauses can also be used as a conversational device to convey uncertainty or to tone down a potentially face-threatening utterance (Levinson, 1983).

2.9.1.3 Prolongations

Prolongation or lengthening is used to describe the phenomenon where a word, or more specifically, a speech sound, usually a continuant (Allwood, Nivre & Ahlsén, 1990, p. 12), is lengthened or is "stretched out ... as a means of hesitation" (Eklund, 2001, p. 5). This is illustrated in the following example from Rose (1998), where the word to is prolonged.

//WELL .// it goes back to: always wanting to be a MISSIONARY //

(n.b. prolonged word in bold followed by ‘:’)

A word is considered to be prolonged if the segment that is being "stretched out" is auditorily longer than the realization of the segment in fluent contexts (Klatt, 1976, pp. 1213-1215). Auditory perception is usually accompanied by acoustic measurements of
the duration of the segment or the entire word that is prolonged (Shriberg, 1995; 1999; 2001). Shriberg (1995; 1999; 2001), for instance, examined lengthening by comparing the duration of a word to its duration in fluent contexts. Similarly, Wightman, Shattuck-Hufnagel, Ostendorf & Price (1992, p. 1711), also used expected duration as a baseline measurement. Following Klatt's (1975) work on vowel lengthening, they looked at segment duration in relation to its expected length. Since some sound segments have a wider range of durations compared to others, they may not necessarily be lengthened even if they are, for example, 30msec longer than the average durational measurement (Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992). Hence, they measured lengthening according to its "normalized duration" by measuring "the duration of a segment as the number of the standard deviations from the mean duration of the phone contained in the segment".

2.9.1.3.1 Categorization of Prolongations

As a hesitation device, prolongations have been found to be a common disfluency after silent and filled pauses (Eklund, 1999, p. 2001). They are similar to filled pauses to the extent that both signal hesitation "by means of vocalization and duration" (Eklund, 2001, p. 5). Shriberg (2001, p. 161) found that prolongations tend to have a similar pitch contour as filled pauses, that is, flat or slowly falling, thus, "allow[ing] speakers to pause without ceasing phonation". However, unlike filled pauses which occur after the interruption point, prolongations occur before this point, and can, therefore, be followed by another form of hesitation or immediately by the repair. Shriberg (2001, p. 161) points out that this suggests that "speakers are signalling delays, hesitating much like they might display with a filled pause".
Because of their similarity with filled pauses, prolongations are also considered planning devices (van Donzel & Koopmans-van Beinum, 1996). In their study on pausing strategies used by Dutch speakers, van Donzel and Koopmans-van Beinum (1996) found two kinds of prolongation, which were used as a means to pause during speech. They are as follows:

(i) the adding of a schwa after the final consonant of a word

(ii) the lengthening of a word-internal vowel or a consonant.

Due to its function as a hesitation device, Hieke (1981) categorized syllabic prolongations as part of stalls in his taxonomy of hesitation phenomena (see Figure 2.3).

Prolongations are commonly found in repeats, where the first token of the repeat or R1 (considered the reparandum) is prolonged compared to the duration of the second token of the repeat or R2 and the duration of the same word in a fluent context (Shriberg, 2001). This corresponds to Den and Clark’s (2000) finding of repeats in Japanese, thus suggesting that such prolongations of the first token of repeats are used as a form of hesitation. Nevertheless, R2s have also been found to be prolonged (Plauché & Shriberg, 1999).

The similarity between the function of filled pauses and prolongations as hesitation devices, affects the way in which the latter is categorized. Faure (1980, p. 287) and Abou Haidar (2001, p. 33), for instance, categorize prolonged syllables as part of filled pauses, although Duez (1993, p. 27) considers prolonged syllables as non-silent pauses together with filled pauses, false starts and repeats. Such categorization suggests that prolongations should be considered part of the editing phase, which occurs after the
interruption point. However, as mentioned earlier in this section, the prolonged word usually precedes the interruption point.

Thus, there could be a problem with placing prolongations within the structure of disfluencies. Since prolongations are said to occur before the interruption point, then they should be part of the reparandum as illustrated in the following examples taken from the data in this study. The case for a prolonged word to be considered as part of the reparandum becomes stronger if it is subsequently substituted or deleted or repeated (see Example 1 and 2). The problem of categorizing prolongations arises in cases where there is no overt error and subsequent repair being made (see Example 3).

Example 1

C2
[I’m calling from:] (DEL) IP actually I’m driving.

Example 2

C9
< [but: ] IP but (REP) > we don’t have the …

Example 3

C14
there’s a lot of cars stopping IP :558 er to: IP have a: IP look.

It appears that in Example 3, the speaker is hesitating by lengthening the vowels in the words to and a. The prolonged words in this example are not ‘errors’ as speech resumes immediately after these words. They, therefore, cannot be considered as part of the reparandum, but should be considered as hesitation markers. Thus, prolongations are a type of hesitation and can also be a part of the reparandum.
2.9.2 Types of Self-Repairs

Apart from hesitation devices like silent pauses, filled pauses and prolongations, by comparing the material in the reparandum with the repair, it can be noted that speakers sometimes

(i) insert new items in the repair,

(ii) substitute items in the repair,

(iii) delete items that were in the reparandum

In fact, insertions, substitutions and deletions are regarded as classes of disfluencies, together with repeats in the the Human Communication Research Centre’s Coding Manual (Lickley, 1998), a classification which has been adopted by other studies such as Aylett (2003) and Shriberg (2001).

2.9.2.1 Repeats

Repeats are a common form of disfluency (Clark & Wasow, 1998, p. 202). Lickley (1998) considers repeats to be instances where “...strings are repeated verbatim with no addition or deletion. The repeated strings may be word-fragments, whole words or sequences of words”. Examples of repetition following the above definition from Lickley (1998, p.1) include:

(i) no [ɪ] straight southwest (fragment)

(ii) right to [mɪ] my left (word)

(iii) right [θɛər's əf] there’s a line about quarter of the way down (sequence of words)
The examples show that repeats should comprise at least two parts, the “repeatable element”, which refers to the first occurrence of the repeat and the “repeated element”, which is the occurrence following the interruption point. (Henry & Pallaud, 2003, p.77). As defined and illustrated in Lickley’s examples, repeats can be “complete”, that is all the words in the repeat are complete words, or they can be “partial” (Tseng, 2003, p.74). The former involves fragments of a word, as illustrated in the Tseng’s (2003, p.74) Mandarin examples:

**Complete Repetition**

* yin wei  yin wei  ta you  jianshen zhongxin
(because because it has a fitness centre)

**Partial Repetition**

* kan dian-  kan dianshi zuijin you xin dianying
(on the tele- on the television, there is a new file recently)

Thus, repeats could consist of the repeatable element being a fragment or a complete word. An item (fragment, word/s) or token can also be repeated more than once. Henry and Pallaud (2003, p. 77) refer to items that are repeated once as “simple repeats” (for example, *le le*), and those that are repeated more than once as “multiple repeats” (for example, *le le le*) in their analysis of repeats in spontaneous spoken French. They found that most of the repeats (93.8%) in their study were simple repeats. They also reported that for multiple repeats, only 5.3% were repeated twice (*le le le*), with the percentage decreasing for more than two repeats. In other words, multiple repeats were not a common occurrence. This is consistent with Shriberg and Stolcke (1998) who found that speakers tend to retrace one word.
When the token preceding the interruption point is a fragment, it can be difficult to
determine if the fragment and the word/s after it is a repeat. This is because since the
whole word is not produced by the speaker, we can only infer, based on auditory and
acoustic similarities between the first segment/s of the token preceding and following
the interruption point, that that the fragment is part of the same word that is repeated.
There are also problems related to determining if we should consider the items before
and after the interruption point as repeats only if complete word or words are clearly
repeated, but not if fragments are involved. Such problems can be seen in Shriberg’s
(1999) examples of four different types of transitions of fragments. The types of
transitions found in a study of repeats of I and the are shown in Table 2.5.

<table>
<thead>
<tr>
<th>Transition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) None</td>
<td>the, the dog</td>
</tr>
<tr>
<td>(b) To word after</td>
<td>the (d) the dog</td>
</tr>
<tr>
<td>repeat having</td>
<td></td>
</tr>
<tr>
<td>(c) To different</td>
<td>the(d) the cat</td>
</tr>
<tr>
<td>word</td>
<td></td>
</tr>
<tr>
<td>(d) To repeat itself</td>
<td>the(h) the dog</td>
</tr>
</tbody>
</table>

(Shriberg, 1999, p. 620)

In the first type of transition (a), the word is repeated verbatim. There is no perceivable
transition but there could be a pause between the token preceding (R1) and following
(R2) the interruption point. In (b), the speaker seems to have committed to a word (i.e.
dog), but appears to be stalling, perhaps, Shirberg suggests, to maintain syntactic or prosodic harmony. The third type of transition (c) is said to be a covert repair, to the extent that the speaker appears to have prearticulatorily detected the problem, causing him to stop at the initial articulation of the “wrong” word, that is, at the segment d, and then proceeds to produce the repair, cat. In the fourth type of transition (d), the fragment th is the initial segment in the previous and following word. Shirberg suggests that the speaker has decided to repeat the word even at the production of R1.

In terms of classifying disfluencies, such fragments could pose a problem. For instance, the fragment in (b) and (d) could be categorized as repeats (Lickley, 1998). However, the fragment in (c) should be considered as a segment that is substituted rather than repeated, since it is the article the and not the fragment d which is repeated. Thus, a decision has to be made as to how such co-articulatory effects are to be treated in the classification of disfluencies. This is especially so when there is no silent or filled pause between the repeats, making it difficult to discern the exact point at which the offset of the first token and the onset of the second token end and begin.

There is also the question of how repeats fit into the structure of repairs since the R1 is not actually corrected, as can be seen in the examples in Table 2.5. Nonetheless, the last repeatable element preceding the interruption point is usually considered the reparandum (see example 1 in 2.7), while the repeated element in regarded as the repair, in the sense that this is the point at which the speaker resumes his speech. This is illustrated Figure 2.8, which shows how repeats fit into the structure of repairs (Henry & Pallaud, 2003, p. 77).
Figure 2.8

Structure of a Repeat

In Figure 2.8, R₀ (note that Shriberg (1995) refers to the first and second tokens as R₁ and R₂ respectively) is the repeatable element, while R₁ is the repeated element. RM is the reparandum, and RR the Repair, while IM is the interregnum, otherwise referred to as the editing phase in the structure of repair.

2.9.2.1 Classification of Repeats

Maclay and Osgood (1959, p. 39) found that repeats tend to involve function words, but similar to filled and silent pauses, occur before lexical words. Thus, repeats could be said to behave like hesitations, “providing time for selection among diverse lexical alternatives” (Maclay & Osgood, 1959, p. 39). Similar findings were reported by Henry and Pallaud (2003). The latter found that 91% of repeated words in their study based on spontaneous spoken French were function words (41.5% determiners, 35.3% pronouns and 13% prepositions). Benkenstein and Simpson (2003), who were looking at spontaneous German speech, also found that most of the words immediately before the interruption point in repeats were function words. In contrast, repeats in Japanese were found to be content rather than function words (Den, Ishizaki & Haruki, 1997, cited in Den, 2001). This is presumably because Japanese has many long content words.
In relation to this, repeats are identified and categorized not just according to the class of words that they are, but also according to their function in the utterance (Dickerson, 1971; Hieke, 1981). For instance, as outlined in 2.5 and Figure 2.3 earlier in this chapter, Hieke (1981, pp. 152-3), divides repeats into “retrospective” and “prospective” repeats. Retrospective repeats or “bridging” act as bridges, connecting speech before and after the hesitation. This is illustrated in the following example taken from Hieke (1981, p. 153), where the second the is said to re-establish fluency which was disrupted by the silent pause:

The cottage was in a place called Harrison /960/ well /3120/ the /1760/ the father in the family ...

(n.b. repeat in bold, numbers indicate time paused)

Accordingly, Hieke categorizes such repeats as “repairs” in his taxonomy of hesitation phenomena. On the other hand, Hieke categorizes prospective repeats as “stalls”, because such repeats are said to act as a stalling device, buying more time for the speaker. This can be seen in the following example (from Hieke, 1981, p. 152), where the speaker appears to repeat to the (in bold) while searching for the next word:

if the cook was /480/ ah /400/ nice an and ah /1200/ to the to the family...

Hieke notes that these two functions of repeats had already been discussed by Dickerson (1971). Dickerson (1971, p. 197 quoted in Hieke, 1981, p. 152) saw repeats as being “multi-functional”:

Its principal function appears to maintain the wholeness of a unit (word or phrase) where that wholeness is threatened, as may happen when pronunciation difficulties and planning pauses interrupt the unit. As a secondary function, however, repetition may also serve as a pause device through habitual usage.
Thus, repeats (e.g. Hieke’s *prospective repeats*) can function in the same way as silent pauses, filled pauses and prolongations, acting as pausing devices (Duez 1993; Postma, Kolk & Povel, 1990). They can also be considered as a type of self-repair which re-establishes fluency, which would explain its classification as a self-repair and also why the repeated element is considered as the repair in such cases.

Shriberg (1995) examined the acoustic features of prospective and retrospective repeats to see if these features could be used to explain the difference between these two types of repeats. This was based on the premise that in prospective repeats R2 is followed by a pause, while in retrospective repeats, R1 is followed by a pause. Shriberg’s (1995) illustration of the surface structure of both these repeats is shown in Figure 2.9.

```
Prospective: R1 (...) R2 ... Continuation
Retrospective: R1 ... R2 Continuation
```

“...” = pause, “( )” = optional

(Shriberg, 1995, p. 384)

**Figure 2.9**

Surface Patterns for Proposed Functions of Repeats

According to Shriberg (1995), in retrospective repeats, R1 is lengthened, while R2 is not, since it is supposed to be the point at which speakers resume their utterances. In contrast, for prospective repeats, there is lengthening of R2 since the repeated token is acting as a stalling device. The R2 in prospective repeats were also found to have falling F0 (fundamental frequency) values similar to filled pauses, thus strengthening claims that such repeats function as stalling devices. Plauché and Shribergs’ (1999) analysis of
the acoustic and prosodic features of repeats, resulted in three sub-classifications of disfluent repeats. Their findings are summarized in Figure 2.10.

(i) Retrospective Repeats

... (Possible Pause) the+++ (Long Pause) the ( - ) ...
Orig. P1 R1 P2 R2 P3 Cont.

(ii) Stalling Repeats

... ( - ) the+ (Possible Pause) the+++ (Possible Pause) ... 
Orig. P1 R2 P2 R2 P3 Cont.

(iii) Covert Self-Repair

... (Often Pause) the+ ( - ) the+ ( - ) ...
- CRK -
Orig. P1 R1 P2 R2 P3 Cont.

Orig. = Original, Cont. = Continuation, + = longer than fluent duration, ( - ) = no pause, Dark lines = pitch pattern, CRK = creaky voice/glottalisation

(Plauché & Shriberg, 1999, p. 1515)

Figure 2.10

Sub-classification of Disfluent Repeats

Figure 2.10 shows that in retrospective repeats, R1 is much longer, and is followed by a pause indicating that the speaker has run into problems (also suggested by the possibility of a pause preceding R1), and is stalling by lengthening R1. There is no lengthening of R2 or pause expected after R2 as fluency is re-established at this point. Plauché and Shriberg (1999) compared the acoustic properties of this repair with covert
self-repairs, where the speaker is said to have detected a problem prior to articulation. In covert self-repairs (see iii in Figure 2.10), there is no pause expected between R1 and R2 in covert repairs, since the problem has presumably been detected, and the repair has been completed by the time R2 is articulated. Both R1 and the R2 have a rising pitch compared to the R1 and the R2 in retrospective repeats. This is because, according to Plaché and Shriberg (1999), in covert self-repairs, R2 “usually marks the beginning of a new utterance or a corrected version of the previous one”. The expected pause before R1, the glottalization on R1 and the rising pitch over R2 suggests that the problem is detected pre-articulatorily while R1 is being produced.

In contrast to retrospective repeats and covert self-repairs, there is no pause preceding R1 in prospective or stalling repeats (see ii in Figure 2.10), although there is a possibility of pauses preceding and following R2. The R2 also has a longer duration than R1. This suggests that a speaker encounters problem during and after the production of R2, but not before R1. There are also said to be the presence of disfluencies after R2 indicating that a speaker is still having problems after the production of R2.

However, both prospective and retrospective repeats are manifestations of covert repairs as defined by Levelt (see 2.8.2) as there is no overt evidence of an error and a corresponding repair. The acoustic features (lengthening, pitch and glottalization) and the presence of silent and filled pauses before and after R1 and R2 may be indicative of the type of possible repair being attempted. Nevertheless, without direct evidence, we can, perhaps, go as far as saying that the speaker seems to be facing a problem. However, what the production-related problem is cannot be determined.
While Plauché and Shriberg (1999) consider only retrospective repeats as instances where “a speaker has not yet committed to following material”, Clark and Wasow (1998) see repeats in general as strategies used by speakers when they have committed themselves to an utterance that is problematic. Repeats are, therefore, seen to be a manifestation of speakers’ inability to continue their utterance, resulting in them hesitating by repeating words with or without any other form of hesitation. The R2 then is seen as their attempt to continue where they left off. In relation to this, in their commit-and-restore model of repeated words, Clark and Wasow (1998, pp. 203-4) divide repeats into the following four stages, which are illustrated in Table 2.6 for the utterance, *I uh I wouldn’t be surprised at that:*

(i) initial commitment

(ii) suspension of speech

(iii) the hiatus (what is between the suspension and the restart)

(iv) restart of constituent
<table>
<thead>
<tr>
<th>Stage</th>
<th>Speaker S's action</th>
<th>Spoken example</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Initial commitment</td>
<td>S commits to a constituent</td>
<td>I</td>
</tr>
<tr>
<td>II Suspension of speech</td>
<td>S stops vocalizing</td>
<td>-</td>
</tr>
<tr>
<td>III Hiatus</td>
<td>S deals with potential delay</td>
<td>uh</td>
</tr>
<tr>
<td>IV Restart of constituent</td>
<td>S restarts the constituent, restoring continuity to it</td>
<td><em>I wouldn't be surprised at that</em></td>
</tr>
</tbody>
</table>

(Clark & Wasow, 1998, pp. 203-4)

In this model, repeats are seen to occur as a result of speakers committing to speak, perhaps when they are not totally ready to proceed. Thus, speech is suspended soon after the initial commitment (in II), as speakers hesitate. Speakers are said to then repeat a word or words after the suspension of speech (that is, I is repeated in IV) "to restore continuity" (Clark & Wasow, 1998, p. 236). This is similar to what Hieke (1981) refers to as retrospective repeats. Thus, repeats appear to be used as a hesitation device in speech, to allow repair-planning (for example, a lexical search) and also as a restart mechanism to maintain fluency in an utterance. Further, as Den (2001, p.28) explains,

*Speakers sometimes use repetitions as a linguistic device to signal to their addressee that they are having trouble in speech production, and produce these tokens [first phonemes of the first token] with the intention of making the addressee recognize them as a signal.*
2.9.2.2 Insertions

In contrast to repeats which involve repeating segments, word or words without any additions or deletions, insertions are regarded as instances when a speaker repeats only some part of his utterance, and inserts a word or words in the repair (Lickley, 1998). Examples of insertions from Lickley (1998) are as follows:

(i)  \textit{[just]} \textbf{I'm just} to the east of it

(ii) \textit{... but [includes] within the line it includes} the concealed hideout

(iii) \textit{just above [a forest fire] site of a forest fire}

(Reparanda in bold and in square brackets; Repair underlined, Insertions in bold)

In the examples above, it appears that words are inserted prior to the repeated items in order to make the speaker's intended meaning clearer, for example by adding more specific information. However, an insertion can also be made within the repeated segments in the repair such as in the following example from Shriberg (1994, p. 14), where the inserted item is shown in bold:

\textit{show the flights -- show the morning flights...}

Thus, Allwood, Nivre and Ahlsén (1990, p. 17) postulate two types of insertions. In the first type (see Lickely's examples), the insertion is made prior to the repeated elements. In the second type of insertion, the insertion is placed in between repeated elements (see Shriberg's example). Whatever the case, the necessary condition is that the insertion has to be placed before repeated elements (Allwood, Nivre & Ahlsén, 1990, p. 20).
2.9.2.3 Substitutions

Similar to insertions, substitutions can also involve some of the words preceding the interruption point being repeated. As shown in the following examples from Lickley (1998), substitutions occur when a speaker replaces a sound, word or words in the repair (Lickley, 1998).

(i) *And you [d-] finish your curve at...*

(ii) *[a vertical] a horizontal line*

(iii) *I don’t suppose you’ve got [the balloons] the baboons*

The examples show the different levels that can be substituted. In example (i), the sound *d* is replaced by *f* resulting in the word *finish*. The problem here is that it is difficult to ascertain whether the substitution is that of a sound or the entire word as we so not know what word the speaker intended to produce based on the fragment. In fact this could even be an instance of a sound deletion.

This contrasts with example (iii) where the whole word is produced, following a phoneme substitution. Here the second consonant /l/ is substituted with /b/ resulting in a different word altogether. In example (ii), the wrong lexical choice seems to have been made, resulting in the substitution of the word *vertical* for its antonym *horizontal*.

Bear, Dowding, Shriberg & Price (1993) even include the substitutions of contractions in their labelling system of self-repairs. This can be seen in the following utterance from Bear, Dowding, Shriberg & Price (1993), where the contracted form *’m* replaces the form *’ll*. 
All right I'll    I'm    interested in flight five eleven

$M_1 \wedge R_1 | M_1 \wedge R_1$

(M = Matching; R = Replacements; numerical value in subscript = corresponding word in the repair; "\wedge" = contracted form)

(Bear, Dowding, Shriberg & Price, 1993, p. 4)

The examples given thus far show a particular characteristic of word substitutions, that is, the substituted word and its substitute tend to be of the same semantic or syntactic category (Fay and Cutler, 1977; Fromkin, 1971; Hotopf, 1980; Raymond & Bell, 1998).

This has implications on the way words or items are stored in the mental lexicon (Levelt, 1989, p. 6) and retrieved during speech production. For example, it has been suggested that this phenomenon shows that items that share similar meanings or belong to the same semantic field are connected to each other in the lexicon (Levelt, 1989, pp. 183-4). It also raises the question of whether items are stored according to their syntactic category although there is no concrete evidence of this.

This characteristic can also help us to identify repairs more accurately. This is especially so in cases where it has to be decided whether an item is being deleted or substituted. According to Bear, Dowding, Shriberg & Price (1993, p. 5), if there is no observable semantic or syntactic relationship between the corresponding words before and after the repair site, the disfluency is categorized as a deletion rather than a substitution. This is illustrated in the following utterance, where *the* (in bold) is considered as a deletion as it does not relate syntactically or semantically to *types of*, which in turn is considered an insertion, thus making it possible for two types of self-repairs to occur within the same utterance.
List the aircraft list types of aircraft.

$$M_1 \ X \ M_2 \ | \ M_1 \ X \ X \ M_2$$

(M = matching; R = Replacement; X = insertion (to the right of bar) or deletion (to the left of bar); | fresh start with semantic relationship before and after the repair site)

(Example from Bear, Dowding, Shriberg & Price, 1993, p. 5)

However, the insertion in the example above does not correspond the one discussed in 2.9.2.2, as the word the to the left of the bar has been deleted. If this word were not present, then the words types of can be considered as an instance of deletion.

2.9.2.4 Deletions

Apart from inserting or substituting items, a speaker can also delete sounds or words from the original utterance. Deletions are, therefore, “words deleted from the RM [reparandum] and [have] no corresponding words in the RR [repair region] ...” (Shriberg, 1994, p. 61, emphasis added). More specifically, deletions are said to occur when the “speaker interrupts their utterance and restarts without repeating or directly substituting any word or structural unit” (Lickley, 1998, emphasis added).

The following examples Lickley (1998) show that the speaker had decided to expunge the part of his utterance in the square brackets, and start anew. They also show that the parts that are deleted can range from a sound segment to a series of words:

(i)  \[s] you then come up north

(ii) well go down [until you're in a ver-] right see the window of the flagship

(iii) [well the bottom of it] right just draw a straight line
2.9.2.4.1 Deletions and False Starts

The examples from Lickley (1998) also show that deleted items can also be part of false starts, where a speaker starts an utterance but then abandons it (Hutchinson & Pereira, 2001), especially when they occur at the beginning of an utterance such as in example (i) and (iii). Thus, the examples of deletions above are essentially the same as the following examples of false starts given by McAllister, Cato-Symonds and Johnson (2001) and Rose (1998), where the items preceding the interruption point (in bold) are deleted and not pursued after that.

Example (i)

*And he's got, he's got a little triangle hat and it - you can see a thick neck and a thick body*

(n.b. false start in bold)

(McAllister, Cato-Symonds & Johnson, 2001, p. 65)

Example (ii)

// *I still have* I'm twenty-SEV en now //

(Rose, 1998)

Similar to deletions, the items that are abandoned are considered as the reparandum, while the new items following the interruption point are considered as the repair (Bear, Dowding, Shriberg & Price, 1993), although how many items after the interruption point make up the repair has to be decided (for example, up to the next error, or hesitation, or next sentence boundary). The problem of relating the repair to the reparandum arises because the entire reparandum is deleted, and thus, it is difficult to establish a relationship between the deleted items and the repair. This is unlike insertions, where part of the reparandum is repeated, or substitutions, where it is possible in most cases to ascertain the item being replaced. This is why only the deleted
item or portion of an utterance is indicated, such as in the previous example from Bear et al (1993).

Perhaps, in an attempt to determine the extent of the repair, utterance-initial deletions are sometimes examined in terms of whether speakers “backed up” or did not back up in order to correct their initial utterance. Maclay and Osgood (1959, p. 24) referred to these as “retraced” and “non-retraced” false starts, respectively. Examples of these false starts from Maclay and Osgood (1959) are as follows:

**Retraced False Start**

*I saw a very big// a very small boy* (self-interrupted utterance)

**Non-Retraced False Start**

*I saw a very ...* (incomplete utterance)  
(Maclay & Osgood, 1959, p. 24)

However, the example of the retraced false start above corresponds to a lexical substitution (*small* for *big*). This is why Shriberg (1994, p.14) maintains that “backing up to a sentence boundary appears to be a feature independent of the type of change effected from RM [reparandum] to RR [repair region]”. She provides the following utterances to show instances where the speaker has retraced to the start of the sentence, and at the same time is involved in making changes to his utterance (by repetition, insertion and substitution).

(i)  show the – **show the** morning flights to boston (repetition)

(ii) show the flights – show the **morning** flights to boston (insertion)
(iii) show the morning — show the **evening** flight to boston

(substitution)

(Shriberg, 1994, p. 15)

Such examples would involve fresh starts, with different kinds of repairs being made, rather than false starts since the speakers back-track or retrace their utterances, in the repair in the former (Shriberg, 1994). This is similar to Rose (1998), who considers **restarts** to be instances when speakers begin an utterance and “then suddenly return to the beginning and iterate the same words” such as in the following example taken from Rose (1998):

```
// but yeah my first r *my first reaction* to THAT - // cm was a reaction
to mySELF //
```

This example is similar to example (i) from Shriberg (1994), as they are both instances of repeats. In other words, repeats, substitutions and insertions can simultaneously be instances of fresh starts or restarts.

However, the following examples (from Shriberg, 1994) are categorized as deletions, since they do not involve retracing to the beginning of the sentence:

(i)  *show me the -- which morning flights go to boston*

(ii) *show the flights that – to boston*  

(Shriberg, 1994, pp. 14-15)

Thus, in order to distinguish restarts or fresh starts from false starts, two parts of an utterance need to be examined: the part in the reparandum where the speaker starts
speaking and then stops, and the part in the repair region, where the speaker may begin a new utterance (fresh start or restart). Utterance-initial deletions can be considered as false starts, while repeats, insertions and substitutions can be considered as restarts or fresh starts if the speaker goes back to his original utterance whilst making the necessary repair.

2.10 Speech Disfluencies and Self-Repair in Second Language Production

Since this study involves the use of English by non-native speakers, a review of studies on disfluencies produced by non-native speakers of a language was done. It was found that most of the studies on speech disfluencies and self-repair in second language (henceforth, L2) production tend to focus on second language learners (e.g. Hieke, 1981; Kormos, 1999; Poulisse, 1999; van Hest, 1996), rather than on competent L2 users, who are able to use the L2 for communicative purposes.

In one of these studies, Hieke (1981) compared the patterns of hesitation among 10 native speakers of German, 10 native speakers of English, and 29 learners of English. Predictably, the non-native L2 learners made more repairs, although all three groups produced almost equal percentages of hesitation or stalls (silent and filled pauses, prolongations and prospective repeats). Raupach (1980, p. 297), on the other hand, found that his subjects had more pauses in their L2 than in their first language (henceforth L1). His subjects comprised German and French undergraduate students, who had to describe cartoons shown to them in both their L1 and L2 (French and German or vice versa).
Similarly, in another study involving the use of hesitation strategies in L2 oral tests by Canadian learners of German, Rieger (2003) found that the strategies used by students were related to their level of proficiency. Thus, students with better grades, like L1 speakers used a variety of hesitation strategies, such as using German lexical fillers and repeats. In contrast, those with lower grades, mainly used unfilled and filled pauses. According to her, this is because “as the knowledge of the target language increases and this knowledge is more efficiently applied, L2 speakers allocate more attention to monitoring their speech performance at the discourse level” (Rieger, 2003, p. 44).

In another study involving the comparison of the use of L1 and L2, van Hest (1996; 1998) found that her subjects also made more repairs in their L2 (English) compared to their L1 (Dutch). They were found to be making more error-repair than appropriate-repairs in their L2 (see 2.8.3). This trend is explained by the fact that learners tend to make more errors since they are still building upon their linguistic knowledge in the L2.

van Hest also compared the level of proficiency of the subjects and their use of self-repairs. The subjects were divided according to their levels of proficiency:

- low (4 years of study, aged between 14-15)
- intermediate (6 years of study, 16-17 years)
- advanced (8 years of study, 20-23 years).

She found that there was no significant difference in the number of repairs produced between the lower and intermediate group of learners. However, there were significant differences in the number of repairs produced between the advanced group and the other two proficiency levels. van Hest suggests that this is because the lower and intermediate
groups are in the process of “trial and error”. van Hest found that the advanced L2 learners produced more “complex” repairs than the other two groups, presumably because they have more linguistic knowledge to draw upon.

Another study that compared Dutch learners of English, Poulisse (1999) studied slips of the tongue made by 45 Dutch learners of English from three different L2 proficiency levels. Consistent with the other studies, the subjects made more slips in their L2 than L1. Similar to van Hest, it was also found that slips decreased with increasing proficiency. Most of the slips made in the L2 production were at the lexical level. There were also many slips at the morphological and phonological levels, and all these slips were found to increase with decreasing proficiency levels (135). According to Poulisse (1999, pp. 136-7), “this suggest that the processes of lexical, morphological and phonological encoding are at different stages of automatization for the three proficiency groups”. In other words, speakers can only monitor their speech and make the appropriate repair according to, for example, the lexical, morphological, and phonological knowledge they possess. Thus, if learners do not have a particular lemma stored in their mental lexicon, or do not have a complete inventory of the lemma’s features (see 2.2 and Figure 2.1), then they would not be able to retrieve the lemma or detect any error related to it.

The general findings from the studies above can be summarized as follows:

- Learners tend to make more repairs in their L2, because they are still in the process of learning the ‘rules’ in that language (Hieke, 1981; Poulisse, 1999; van Hest, 1996, 1998).
• Accordingly, the more proficient an L2 learner is, the less errors he tends to make (Poulisse, 1999).

• Consequently, since he makes less errors, the more proficient or advanced L2 learner will probably produce less repairs than the less proficient one (Hieke, 1981; van Hest, 1996, 1998).

• However, the more proficient learner can be expected to make more complex repairs since he would have more linguistic knowledge of the L2 (1996; 1998).

• L2 learners tend to pause or stall more often in their L2 than in their L1 (Raupach, 1980, p. 297), with more proficient learners tending to use more complex fillers similar to ones used by L1 speakers (Rieger, 2003).

Because there is undeniably a link between the level of proficiency in an L2 and the production of errors, hesitation and self repair, studies have tended to focus on comparisons of disfluencies produced in L1 and particularly L2 learners. However, there is hardly any research that treats L2 speakers as actual users rather than learners of the language. By treating L2 speakers as users, whose main aim would be to communicate their message, more emphasis can be given to an investigation into the patterns of speech disfluencies in normal spontaneous conversations conducted in L2. The speakers may be bilingual or multilingual speakers who have the L2 as one of the languages in their linguistic repertoire. This would move away from regarding disfluencies as a measure of low proficiency among L2 speakers. As Chambers (1997, p. 538) rightly points out,
In verbal encounters common to language users, mainly conversations and discussion, pauses and hesitations are normal features of the interaction. Despite this, silence is often seen as a sign of disfluency, especially in foreign language speech where it may be perceived as signalling poor functioning of mental processes, instead of viewing it as a normal feature of speech processing.

Thus, disfluencies, such as silent and filled pauses, prolongations, repeats, insertions, substitutions and deletions, are seen as naturally-occurring phenomena in speech, which are part of the normal process of producing speech.

2.11 Chapter Summary

This chapter showed how the many aspects of speech disfluencies, such as slips of the tongue, speech errors, and hesitation and self-repair, and the different types of speech disfluencies, such as silent and filled pauses, prolongations, repeats, insertions, substitutions and deletions relate to each other in the context of speech-monitoring, error-detection, the interruption of speech and self-repair. The literature reviewed in this chapter illustrates some of the problems with definition of terms, conceptualization of types of speech disfluencies and methodological issues related to the study of different aspects of disfluencies. The next chapter will focus on the theoretical understanding of the processes of self-monitoring and self-repair in speech production.