Grammar, Prosody and Speech Disfluencies in Spoken Dialogues

Shu-Chuan Tseng
Department of Linguistics and Literature
University of Bielefeld

8th April 1999

to my parents

Contents

List of Figures

List of Tables

Acknowledgments

I am sincerely grateful to my advisors, Prof. Dafydd Gibbon and Prof. Dieter Metzing, for their encouragement and support over the past years. They helped me to realize how it is to be a modern linguist. I have been lucky to be associated with the Graduiertenkolleg, "Aufgaben-orientierten Kommunikation", which allowed me the opportunity to attend numbers of fruitful talks and discussions.

I would also like to thank those in the department who provided me their technical support, especially Max for his advice on statistics, and the spectrum groups in C6 for their listening to my "Irregularitäten". I am also grateful to Ulrich Schade for his constructive comments with respect to speech repairs.

I would also like to acknowledge the financial support of the Nordrhein-Westfalen Graduiertenförderung and the German Research Council (Deutsche Forschungsgemeinschaft).

Lastly, I am happy that I have the support of my family.

Chapter 1

Introduction

Without complete linguistic knowledge of different levels, phonetic, lexical, syntactic, semantic and pragmatic, one is not well-equipped to produce language properly in any form, independent of writing or speaking. From the opposite perspective, even if one is well-equipped with perfect and trained speaking skills, speech errors, hesitations and speech repairs are unavoidable in the production of spoken language. This kind of phenomenon apparently distinguishes spoken language from other forms of language production such as read speech or written language. The use of written texts, read speech and spoken language and their realization have been very frequently investigated by means of psycholinguistic experiments and natural language processing and modelling. This sort of research aims to look for cues for human language processing in all disciplines of theoretical and applied linguistics, through the insight into different types of language production and the related behaviour of the speakers. Spoken language production, as compared to research done in the field of written language and read speech, has not been given much attention to until very recently. Language used in written form usually contains compact and logically structured content, is grammatically correct, seldom includes speech errors and can be repeatedly read. This is seldom the case concerning spoken language. The spontaneously produced speech contains less compact information, is often grammatically incorrect and full of disfluencies. But is it possible for speakers to utter completely grammatically correct sentences just like those in written texts? Is it useful for listeners to perceive and filter correct and important information out of entire discourses, if speakers speak this way? Investigating syntactic violation in the production of spoken language can provide insight into the relationship between linguistic competence and performance of speakers. Those syntactic deviations are a kind of disfluency. Disfluencies such as pauses, repetitions and restarts are very often produced in spoken language. In addition to the syntactic aspects, the use of prosody is another peculiarity of spoken language. Although the importance of prosodic aspects has repeatedly been mentioned in the literature, research focusing on the acoustic-prosodic properties of disfluencies can seldom be found.

This thesis is concerned with the following questions:
1) How are speech disfluencies produced from the linguistic point of view, focusing especially on the syntactic and prosodic features ?
2) Do regular internal relations exist within speech disfluencies and if so, what do they look like ?

This thesis aims to answer these two questions by several empirical studies and by structurally describing and modelling the results of the studies. With respect to the first question, corpus analysis and experimental phonetics are adopted as methods for examining real speech data concerning both syntactic and prosodic characterizations of disfluencies. More specifically, the aim of the analyses is to examine the performative functions of syntax and prosody in spoken language especially from the viewpoint of how (prosodically) and where (syntactically) speakers make errors or repair their speech. It is proposed in this thesis that a certain regularity of syntactic and prosodic properties within and around disfluent speech can be empirically found. To answer the second question the results obtained in the empirical analyses provide clues about what disfluency relations should look like and therefore, how they can be formally described and modelled. Thus, disfluency models and finite state automata concerning the internal structure of disfluencies are proposed to illustrate the linguistic tendency in the form of a regularity in the production of disfluencies.

1.1 Syntax, Prosody and Speaking

When speaking takes place, prosodic patterns can be observed on all linguistic levels such as lexical, morphological, syntactic and semantic stages. Lexical prosody may influence the realization of utterance prosody which again has to do with the surface structure of utterances in respect of their syntactic constructions. The use of prosody sometimes overrides the original function of lexis and syntax, as one can often observe in spoken conversations. As shown in Figure 1.1, each linguistic level has to be realized and combined with corresponding prosodic forms. Take lexical speech errors as an example. The associated prosodic realization of and around an error may be a signal of disfluency and therefore marks the occurrence of this error. In this case, the effect on the linguistic processing of the utterance containing this error shows first at the lexical level projected onto the prosodic platform. The prosodic effect caused by the error persists through all linguistic levels. The utterance prosody is thus influenced by the lexical prosody. From the point of view of the influence of prosody along the horizontal direction, if a word in an utterance is wrong, then accordingly the syntactic surface structure may be wrong as well and probably even the meaning of the utterance.

As illustrated in Table 1.1, linguistic features of written texts can be clearly distinguished from those of spoken language. Temporal features play a central role in distinguishing written from spoken language. The linear distribution of written texts gives readers the opportunity to follow the contents of the discourse step by step. The author attracts the attention of the reader through a spatially organized layout. Except in certain types of elec-

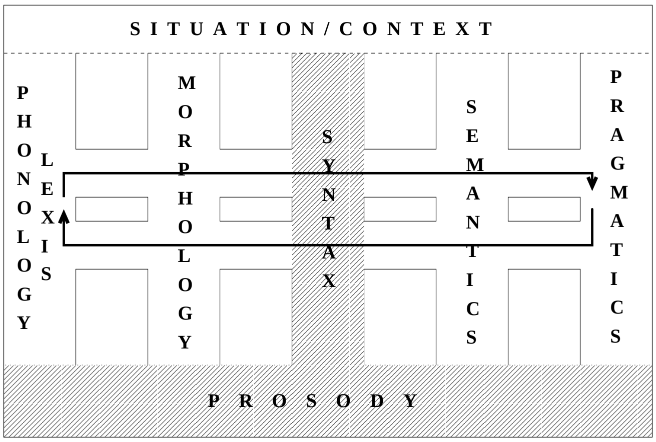

Figure 1.1: Linguistic Levels

tronic communication, the readers have no influence on the communication between themselves and the author, whereas interlocutors in a speaking community are able to influence the communication at any time and the form of influences is strongly speaker and situation dependent. The perception of discourse contents by the addressee is also performed differently in written and spoken language; they are realized in visual and auditory ways, respectively. In the case of a speaking community, the consequence of interpreting speech wrongly is very often quickly shown by false answers or inappropriate reactions. This interaction can also take place simultaneously in the case of spontaneous conversation. In contrast, when the readers read a text there is merely a passive interactive effect between the readers and the text producer. Furthermore, interruptions, different from syntactic pauses made at clause or phrasal boundaries by professional speakers, occur very often in spoken language, whereby in written texts interruptions are mostly used to emphasize specifically chosen phatic functions in conversations. Another important difference between written texts and spoken language which can be observed in the surface structure of utterances is the well-formedness of the utterances according to a given grammar of a language. In written texts, the authors try to produce grammatically correct sentences; the same applies to

Table 1.1: Written Texts and Spoken language

	written texts	**spoken language**
space-time	spatial	temporal
interactive	no simultaneous interaction	simultaneous interaction
communicative	seldom interruptions	frequent interruption
grammatical	mostly well-formed	often ill-formed
prosodic	not prosodically marked	prosodically marked

the case of read speech. In spoken language, however, the speakers on the one hand follow the grammatical rules, but on the other hand, the communicative situation in which they are, including acoustic transmission, visual contact, auditory perception and intention of the interlocutors etc., is the most important factor influencing their speech production. It seems that the importance of grammaticality is strongly reduced in spoken language. It will be shown in this thesis that this is in fact not the case. Syntax still provides a solid linguistic background, even when ungrammatical strings are produced. The use of prosody also distinguishes written language from spoken language. Intention, emphasis, as well as emotion can be directly expressed by means of prosodic markings in spoken language.

Therefore, the use of syntax and prosody will both be investigated in the empirical studies as well as in the disfluency models of disfluency relations. This thesis principally concentrates on the shaded regions in Figure 1.1, focusing on the investigation of particular disfluent phenomena, that is to say, to investigate cases where something goes wrong while speaking. Syntactic features play the most central role in building the construction of utterances, because the levels of phonology, lexis and morphology meet the higher levels of semantics and pragmatics at the interface with syntax.

1.2 Disfluent Speech

Disfluent speech can appear in the form of inappropriately articulated phonemes, false selected lexical items, wrongly inflected forms of morphemes, as well as interrupted melodic realizations of the message delivery. All these phenomena show an effect on syntactic well-formedness. Inappropriate lexical choices, false pronunciations, false syntactic structures and restarts of utterances in spoken language are usually considered irregular. The reason why they are regarded as anomalous phenomena is that they do not fit the given sentence constructions which have been developed for the written language. However, these anomalous phenomena characterize the realization of human verbal communication and distinguish the spoken language most clearly from other forms of language use. One obviously needs a kind of *spoken grammar* specifically for spoken language, instead of applying the written language standard to a very different form of language realization. The following examples of disfluent phenomena in spoken language quoted from various corpus collections may at first sight seem to be anomalous in the transcribed written form, but in fact they can be heard quite often in everyday conversation. Example 1.1 is a restart. The speaker interrupts the original utterance and begins with a new utterance. Example 1.2 and 1.3 are two speech repairs, in which two different speech errors occur, phonological and syntactic errors, respectively. Example 1.4 and 1.5 are speech repetitions with/without editing terms within the repetitions.

Example 1.1 (filled pause, restart)
The girl who said she liked <u>oh</u> <u>Vivaldi dries the dishes</u>. (James 1973)

Example 1.2 (speech repair)
<u>Een eenheed eenheid</u> vanuit de gele stip. (Levelt 1983)
(A unut, unit from the yellow dot.)

Example 1.3 (speech repair)
ähm jetzt das gleiche nochmal <u>für das für die andere gelbe Schraube</u>. (Sagerer et al. 1994)

(Ehm now the same one more time for the[neuter][1]*, for the[feminine] another yellow screw.)*

Example 1.4 (speech repetition)
En aan de rechterkant een oranje stip oranje stip. (Levelt 1983)
(And at the right-side an orange dot, orange dot.)

Example 1.5 (filled pause, speech repetition)
and uh and I've made a career of it. (Blackmer/Mitton 1991)

Researchers in pragmatics and discourse analysis had noticed early on the importance of the structural characteristics in spoken language with respect to disfluent phenomena: [46]; [47]; [48]; [104]; [16]; [66]; [118]; [50]. Pragmatic as well as poetic functions of interjections, speech errors and repetitions were examined, based on practical examples from everyday talk. Some researchers also took other perspectives into consideration. Levelt & Cutler [69] not only collected and classified speech errors and repairs, but also tried to explore the prosodic markings of speech errors and repairs in terms of perceptual examinations, while Levelt [68] furthermore analysed speech errors and repairs from the perspective of speech production. According to Levelt & Cutler [69], utterances which contained erroneous repairs were more frequently prosodically marked, i.e. intonationally marked, than those containing appropriateness-repairs. However, this effect was not found by Nakatani & Hirschberg [82], by means of an acoustical analysis. While Nakatani & Hirschberg [83] used the speech-first model to detect repairs, Bear *et al.* [4] suggested making use of multiple knowledge sources such as pattern matching, syntactic information and acoustic cues to detect and correct repairs.

Hesitation phenomena are usually disfluent and appear in the form of pauses or repeats. In repair research, unfilled pauses and repeats have been considered to be covert repairs. The evidence provided by Postma *et al.* [92] clearly supports the claim that pauses and repetitions have the properties of

[1] Detailed information about this notation cf. Appendix B.

self-repairs rather than of speech errors. Pauses and repetitions are thus "by-products of covert repairing of the internal speech errors". This hypothesis is called the *covert repair hypothesis* in the literature. Are simple disfluent interruptions self-repairs or not? An interrupted syllable is both a speech error and a disfluency. Will it not be more advantageous to have a definition given by the availability of disfluency, instead of a definition which is based partly on the surface form and partly on the communicative functions? After all, some features are shared by speech errors, hesitations and self-repairs. They all occur only in spoken language, vary temporally within discourse and are all disfluent in some sense. Furthermore, terms such as "repair" are to a certain extent ambiguous. For example, the correction within a speech repair is sometimes also called "repair".

In written texts, discourse is segmented by punctuation whereas spoken language uses other means (e.g. discourse particles and prosody) to make the discourse structure explicit. Prosody is one of the important means to explicitness in spoken language and will be investigated with respect to the production of speech disfluencies. Intonation contours, accents, phrase tones and pauses all fulfil the functions of segmenting and emphasizing discourse. Especially in the case of spoken language, interruptions and hesitations are presented by prosodic means, such as interrupted F_0-contours, faster speaking tempo and unfilled/filled pauses. Acoustic-prosodic results can be obtained by means of quantitative measurements and when appropriately modelled can give clear indication of the forms of speech production. Systematic investigations of prosodic features of disfluent phenomena have been seldom carried out, although prosodic information provides a broader perspective for the observation of the speech production process. In automatic language processing, prosodic information is mostly used as an additional parameter used to improve the results [4]; [82]. In psycholinguistic research, psycholinguists try experimentally to predict certain influences of prosody on speech perception. In the present context, prosody is a central concern.

1.3 Related Fields and Applications

In terms of observations and classifications of disfluent phenomena in spoken language, applications can be found in various linguistic research fields by taking into account disfluent speech. These related research fields are 1) automatic speech processing, 2) psycholinguistics and 3) sociolinguistics.

About 10% of sentences longer than nine words in the human-machine corpus studied by Bear *et al.* [4] contained repairs, whereas in Levelt's study [66] 959 repairs were found in 2809 descriptions, about 34% of the entire corpus. In the corpus investigated by Blackmer & Mitton [5] repairs occurred every 4,8 seconds, whereas Lickley & Bard [70] found disfluency on average of every seven words. The task to detect and correct repairs in spoken language, has been undertaken by Heeman & Allen [40]; Nakatani & Hirschberg [82] and Shriberg *et al.* [112]. This high frequency of disfluencies in spoken speech data indicates that one has to deal with disfluencies in order to achieve better recognition results. The application of research on the disfluent discourse structure in spoken language can help achieve the aim of a speech recognizer which understands "normal" human speech. A variety of linguistic information such as syntactic, semantic and prosodic information have been applied to detect disfluency, but rather separately. That is to say, very few of the approaches mentioned above have focused on the combinatory information delivered by more than one linguistic resource.

Analyzing discourse structure of spoken language offers information about the strategies people use when they are hesitating or thinking about how to continue talking. They give clues about how linguistic knowledge is stored, such as syntactic category, semantic meaning, phonological distribution and phonetic shapes of lexical entries. Experiments have been carried out to find out in which way human beings understand the erroneous spoken language with which they are confronted in everyday conversations. Lickley *et al.* [71] suggested with evidence from their experiment that the function of an editing signal as an indication of the impending appearance of repairs was not as significant as suggested by Hindle [42], but that the entire prosodic impression of sentences could play a more significant role [70]. The analysis of spoken language provides a psycholinguistic explanation of how human beings process and understand their language.

Besides, the pragmatic functions of the discourse structure of disfluent spoken language, it also explains the automaticity and prepatternedness of language use [118]. Likewise, the analysis of disfluencies can give information about what the interaction and social relationships of participating interlocutors look like, for instance in terms of a turn-taking system of conversation [96]. The investigation of these essential features of spoken language can shed light on the question of how human beings use disfluencies to keep the floor, to attract the attention of an audience and to make their contribution more understandable.

The results of this thesis can contribute to these three related research fields. By focusing on syntactic and prosodic properties of disfluencies simultaneously, more information about how speech disfluencies are produced can be provided. Thus, this can be adopted for detecting disfluency in automatic speech recognition systems as well as providing stimuli in experimental psycholinguistic research.

1.4 Overview of the Thesis

The structure of this thesis is illustrated in Figure 1.2. The contents of this thesis are summarized chapter by chapter, as follows.

Chapter 2 provides an overview of the state of the art of the disfluency research: how disfluencies have been classified and examined under various viewpoints. It also summarizes research results with respect to the structural and prosodic information provided by disfluencies.

Chapter 3 gives the reasons why the category *phrase* is an appropriate unit for the structural description of disfluencies. It starts with the definitions of terms used in the following syntactic analysis: disfluency-phrase and length of disfluencies. Results of the empirical study show that the majority of disfluencies are produced in constructions with certain syntactic features. The length and category of disfluencies are examined. Subsequently, investigations of the location of interruption and the onset as well as the offset of corrected sequences support the notion that phrases do play an important role in the production of disfluencies, in particular when they are repaired. Sequences such as (*problem phrase, editing phase, corrected phrase*) are very often adopted by the speakers to repair their erroneous or inappropriate speech.

Chapter 4 introduces the results of the acoustic-prosodic analysis of a selected type of disfluency, examined in the syntactic analysis done in Chapter 3. The most common forms of disfluencies were found to be (*problem phrase, editing phase, corrected phrase*). Data produced by a selected female subject in the corpus investigated in Chapter 3 are examined. Following the theoretical backgrounds of the prosodic features which are to be examined, related statistical setting and data processing are introduced. The results show that where syntactic features play a role in the production of disfluencies, significant prosodic effects can also be found.

Chapter 5 is concerned with a re-analysis of the syntactic features of the disfluencies investigated in the acoustic-prosodic analysis in Chapter 4. It is proposed that the results found in the corpus analysis can be re-found in the data produced by the selected subject. This re-analysis aims to secure the empirical support that the syntactic and prosodic cues can be applied to the same data.

Chapter 6 deals with a proposal for a syntax-oriented formal description of speech disfluencies found in the specific corpus. Focusing on prepositional and noun phrases, one has a clear overview of the internal phrasal structure of complex speech disfluencies. It is thus advantageous to extend a description system for complex speech disfluencies to simple ones. Additional prosodic information is used to help disambiguate similar syntactic constructions from speech disfluencies. Notations and formalisms are subsequently introduced to formally describe disfluencies. Furthermore, finite state automata are developed to model the specific characterizations of complex disfluencies, namely the retracing of the repair, the interruption of speech flow and the completion of the repair. The coverage of the occurrences produced by extending the finite state models is also discussed.

Chapter 7 is concerned with specific characteristics of spoken dialogues. Syntax, prosody, interaction, as well as disfluency markers, contribute to the structuring of spoken dialogues. The relationships between these factors are discussed.

Chapter 8 summarizes and concludes this thesis.

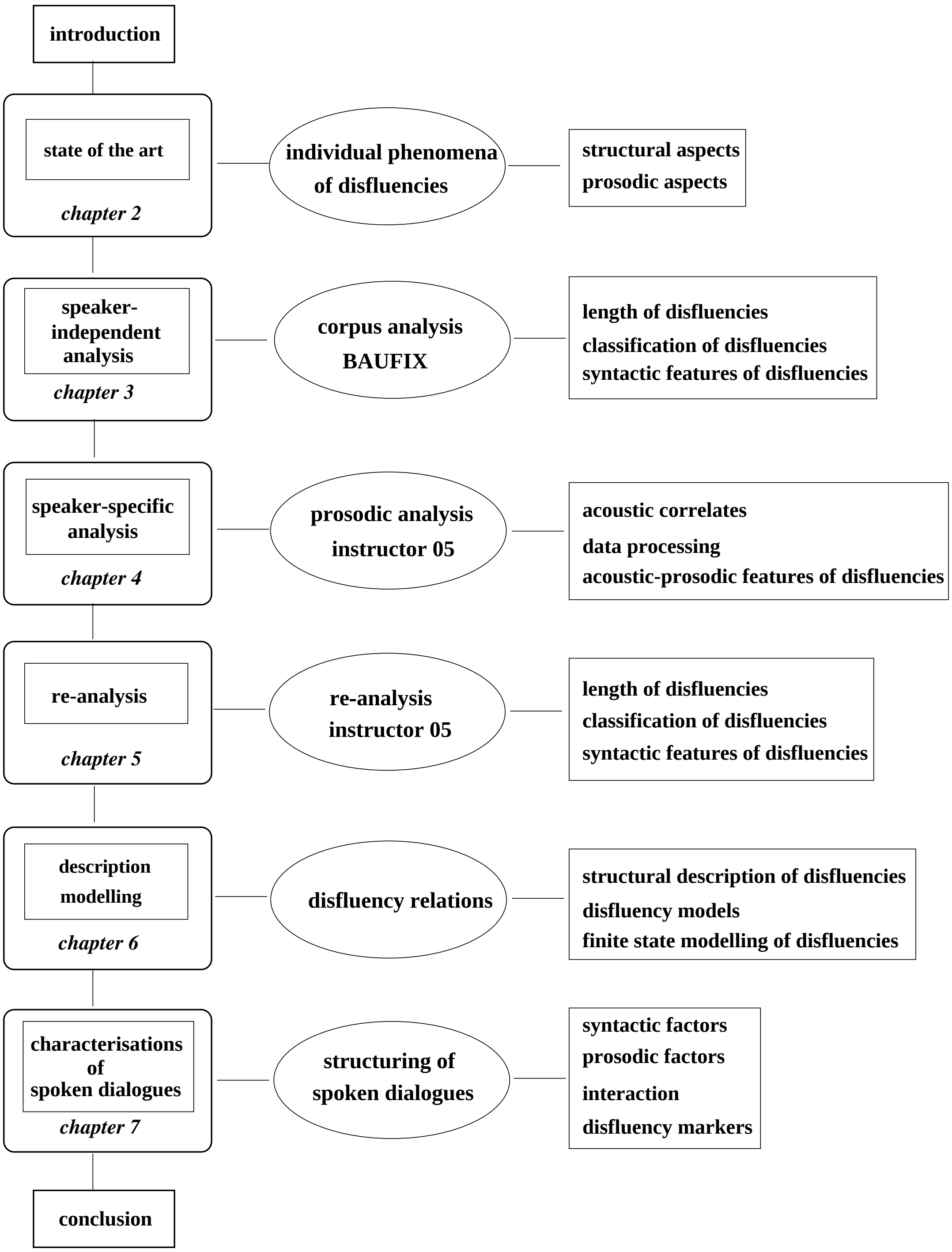

Figure 1.2: Structure of this Thesis

Chapter 2

The State of the Art: Disfluencies in Spoken Language

Simple speech disfluencies include *restarts* (cf. Example 1.1) [75]; [88], *repetitions* (cf. Example 1.4; 1.5) [124]; [111], *silent pauses* [11]; [79] and *filled pauses* (cf. Example 1.1; 1.5) [46]; [48], whereas more complex speech disfluencies are *speech errors* (cf. Example 1.2; 1.3) [29]; [30]; [16]; [108] and *speech repairs* (cf. Example 1.2; 1.3) [66]; [69]; [4]; [39]. A number of linguistic research fields have recently focused on the production of disfluencies in spoken language, from psycholinguistic experiments, concerning language production and perception, to language modelling with applications in the field of automatic speech recognition and natural language processing. Table 2.1 summarizes the disfluency phenomena in spoken language and lists the related research results with regard to disfluencies carried out in recent experiments. This chapter is mainly concerned with these research fields listed in Table 2.1. The disfluency phenomena in different linguistic research fields and applications are to be introduced briefly in the following sections.

Except on particular occasions, sequences of this kind are seldom used

Table 2.1: Disfluencies and Research Fields

disfluencies	**related linguistic research fields**
restarts	language production
repetitions	language perception
pauses	language modelling
speech errors	automatic speech recognition
speech repairs	parsing spoken language

in written language. Unlike written texts and read speech, the discourse structure of spoken language is to a certain extent non-linear and it can in particular vary temporally during the flow of speech. A variety of linguistic occurrences and production forms can therefore be observed in spoken language. Disfluencies are the phenomena which most clearly distinguish spoken language from other forms of language realization. Disfluent speech itself provides important clues to language processing. By investigating the production of spoken language, evidence can be obtained about how speech processing functions when the flow of speech is interrupted and re-initiated [68]. Likewise, erroneous and problematic sequences can reflect how language is processed and repaired by looking at the disfluent phenomena.

Issues in language perception can also be explored by examining disfluencies. To what extent can human beings deal with a more or less large number of disfluencies in everyday conversations and hardly have problems with them ? How disfluencies in spoken language systems are processed has recently been investigated as well. A structural regularity in the production of disfluencies is thus substantially important, for the reason that linguistic cues, especially if available, can mark and signal the occurrence of disfluencies and therefore strongly support the task of detecting and correcting disfluencies in natural language systems. Syntactic and prosodic cues for disfluencies, in particular, play an important role. Syntactic expectancy, based on the grammatical competence the speaker has, is reflected in the production of speech repairs. Why must errors be corrected? In some cases because the content is not exactly correct, while in other cases, syntactical rules are violated and must be corrected. In this thesis, the question is not why errors are corrected, but *how they are corrected*. Therefore, the reflection of disfluencies on the syntactic level is relevant to the structure and the environment of disfluencies. Prosodic cues for disfluencies are also useful in this case in recognizing the existence of disfluencies and subsequently the internal structure of disfluencies. Applying prosodic cues in automatic speech recognition may be helpful in detecting disfluencies before parsing is initiated. Similarly, prosodic cues can provide additional indications for disambiguating syntactic ambiguous strings.

This chapter on the state of the art is divided into three parts: 1) introductory illustration and explanation of disfluencies, 2) summary of recent research results concerning the syntactic-structural features of disfluencies and 3) summary of recent investigations dealing with the prosodic aspects of disfluencies. Terms used for the individual phenomena of disfluencies: *restarts*, *repetitions* and *pauses* as well as *speech errors* and *speech repairs* are to be explained first and subsequently illustrated by concrete examples. Following the clarification of terms and phenomena, the results of research investigations carried out recently in the field of disfluencies are to be summarized, focusing on the structural and prosodic aspects of disfluencies. Finally a general discussion on the relationship of the state of the art and the goals of this thesis will be presented.

2.1 Disfluencies

Research on simple speech disfluencies such as pauses, produced in spoken language, were initially taken up by psycholinguists and conversation analysts. The location and functions of pauses in spoken discourse have been discussed from the point of view of syntax and pragmatics [34]; [9]; [103]; [79]; [60]; [21]; [22]. Similarly, repetitions and restarts have also been dealt with regarding mainly their pragmatic functions [53]; [50]; [119]; [75]; [124]. The tendency of linguistic interest shows that automatic speech recognition of the location of disfluencies [115] as well as related acoustic-prosodic cues and perception experiments are issues which are more frequently focused on. Furthermore, cross-linguistic investigation has also been carried out with respect to the linguistic behaviour of the production of disfluencies [24]. Recently, three dissertations have dealt with speech disfluencies from various analyses. This shows the development of research interest on simple disfluencies in different linguistic fields.

Shriberg [110] gave an overall description of speech disfluencies by investigating three large speech corpora, ATIS [78], AMEX [55] and SWBD [33], in terms of different structural aspects and the related acoustic cues. Results obtained by means of the corpus analysis were afterwards applied to improve the recognition results of automatic speech processing [4]. In contrast to Shriberg who focused more on the production side of disfluencies, Lickley [73] tried to find out by which means disfluencies are detected, by carrying out a series of experiments. The experiments and investigations carried out in his thesis aimed to answer two essential questions: how soon can the listener detect disfluency and what cues can the listener use in detecting disfluency. Subjects could detect disfluency within the first word in the repair, i.e. before subjects had lexical access to the repair. Fox Tree [123] also ran experiments concerning speech perception, but placed more stress on the relationship between speech comprehension and disfluencies. In repetitions such as *the gift the gift department*, the second occurrence of words helped the listeners recognize the missing information; in this case *department*, whereas no recognition benefit was found in repetitions without missing information; that is, in repetitions which only contain simple word repeats. When false starts were located in utterance-initial positions and thus new utterances were begun, comprehension was not slowed. In contrast, comprehension was slowed in the case of utterance-internal false starts; namely speech repairs, because the listeners had to figure out where the false starts began and then where they should attach the restarted information.

When speech errors or repairs occur, the utterances are usually grammatically incorrect, with regard to the notion of *grammaticality* which is developed for written language. Repairs can be initiated and carried out either by the speaker or by the listener, depending on the related communication situation. In this thesis, only self-repairs which are realized immediately after the related errors are made, are taken into account [66]; [104]. A classification of speech repairs usually contains an indirect classification of speech errors. Later in this section, in addition to three significant classification systems of speech repairs, one classification of speech errors is to be introduced.

In the following section, individual phenomena of disfluencies will be introduced in detail, as well as important research results which are directly relevant to the analyses discussed later in this thesis.

2.1.1 Restarts

A restart is when an utterance is interrupted and the speaker begins with a new utterance without finishing or repairing the previously interrupted phrase. *False starts* are the unfinished utterance fragments, for example *you get* in Example 2.1. The new beginning of utterances is called the *fresh start*, for example *it is* in Example 2.1. The whole sequence of fresh starts replacing false starts is called *restart*. If the fresh start is not located at the beginning of an utterance, it is not a restart according to the definition above, but a speech repair.

Example 2.1
I think that you get - it is more strict in Catholic schools. (Hindle 1983)

Detection of Restarts Restarts can be detected in automatic speech recognition, applying temporal features and lexically marked sequences. O'Shaughnessy [87] made use of the temporal intervals of pauses between false and fresh starts to detect the position of restarts. The moment of interruption has proved to be a useful cue to signalling and detecting restarts. For instance, based on these signals the stack copy editor and the lexically triggered restarts are effectively processed by Hindle [42], by comparing the surface form and the related syntactic categories of speech strings (details cf. Section 2.2.3). Comparison of the syntactic categories of phrases and lexical entries which signal a change of intention by the speaker provide efficient criteria for detecting restarts through the surface structure of speech stretches. Interrupted or disfluent intonational or other prosodic cues also indicate the occurrence of restarts. This kind of prosodic cue can be useful in helping listeners detect restarts, as investigated by Local [75]. Local made a descriptive investigation of the phonological-prosodic features of continuing and restarting problems in conversations, aiming to develop a phonology for spoken conversations. Falling-rising pitch contours at the end of fresh starts, i.e. the interruption before the speech flow is re-initiated, is one of the relevant characteristics found in the description. The pitch contour at the beginning of the fresh starts matches roughly with that of the end of falling-rising phrasal fragments. Regarding temporal features, fresh starts, in some cases correction within repairs, are realized significantly faster than the false starts and the reparandum respectively.

Perception of Restarts The effect of restarts on speech comprehension was examined by Fox Tree [124], leading to the conclusion that this kind of speech disfluency does not necessarily slow speech comprehension in spoken language communication, as predicted by language production models from the viewpoint of syntactic theory. Experimental evidence shows that the repair process for restarts is particularly time-consuming in mid-sentence position. In other cases, restarts do not turn out to influence speech comprehension to a significant extent.

2.1.2 Repetitions

Repetitions which have rhetorical purposes and are intended to intensify the effect of an expression, as discussed in [118]; [51], will not be considered in this thesis. Only those repetitions in spoken language are considered which signal the speaker's monitoring of his speech and his problem of continuing speaking. Ochs [86] claimed that repetitions of lexical items are common in unplanned talk. The speaker is trying to gain time by repeating words he has said. Hieke [41] suggested that, with respect to hesitation phenomena, there are two functions which repetitions can serve: stalling and repairing. Repetitions used for simply bridging the thinking phase of the speaker are of the stalling type, called prospective repeats. A retrospective repeat in a repairing repetition has contrastively the property of correcting the repeated item and is not merely a repeat. In [41], this difference of functions was observed through the location of pauses. Pauses are usually located after the repeated words in prospective repeats, whereas they are produced before the repeated words in the case of retrospective repeats. Moreover, this proposal has been approved by Shriberg [111] from the empirical evidence obtained by means of an acoustic analysis. The difference in the location of pauses was approved and there appeared to be a tendency that F_0-reset was more frequently observed in retrospective repeats than in prospective ones.

Postma *et al.* [92] proposed the *covert repair hypothesis*, arguing that repetitions have the property of speech repairs. In fact, in many classifications of speech repairs, repetitions are classified into the group *covert repairs*. There exists another aspect concerning the relationship between repetitions and repairs. That repetitions are indicative of repairs is mentioned by Shimanoff & Brunak [109] and Lickley *et al.* [71]. In the opinion of Shimanoff & Brunak, redundant repetitions function like pauses in signalling grouping. The speaker experiences a temporary memory loss and needs time to get the right words to overcome his lexis-selection problem, in which he simply repeats the word he is saying. In [124] it is shown that repetitions do not have an effect on slowing down comprehension in spoken language. The repair process is inactive while the speaker repeats himself. This supports the redundant property of repetitions which has already been mentioned.

2.1.3 Pauses

Shimanoff & Brunak [109] distinguished two classes of pauses in spoken language: filled pauses and silent pauses. Filled pauses are also called interjections such as *uh*, *um*, *er* in English and *ähm*, *öhm*, *mhmhm* in German. Silent pauses are simply silence with no sounds articulated. Interjections are usually considered non-lexical items, whereas Fischer & Johanntokrax [27] proposed that interjections are available and existent in the lexicon of a competent speaker.

Automatic Speech Recognition It is not long since the acoustic features of pauses have been investigated with respect to automatic speech recognition. O'Shaughnessy [87] for instance made use of prosodic cues of pauses in spoken language to identify syntactic boundaries. In his study, fundamental frequency (F_0) tended to fall prior to an ungrammatical pause and to rise prior to a grammatical pause[1]. From another point of view, in the chapter of acoustic-prosodic analysis, the pitch information before and after pauses will be compared to see whether the pitch contour is continued after a pause has been made.

Use of Filled Pauses Shourup [106] gave a summary of pragmatic functions of interjections in English. He identified seven interjections in spo-

[1] Ungrammatical pauses are pauses which occur within minor syntactic phrases, whereas grammatical pauses are found at the major syntactic boundaries.

ken English: oh1, oh2, oh3, hey, aha, ah and uh[2]. Related to repairs, Schiffrin [105] distinguished the interjections *oh* and *uh* according to their pragmatic functions: *oh* can be used as an indicator of initiating repairs, whereas *uh* is often used as an indicator of continuing repairs. Interjections also indicate that the interlocutors should carry on or yield the floor, i.e. interjections structure the interaction among the communicators. Schegloff [103] for instance investigated the usage of *uh huh* as continuers in respect of the interactional achievement of discourse.

Interjections provide a kind of test of syntactic phrasing, in addition to pragmatic conclusions such as in Example 2.2, mentioned by James [46]. Under normal communication circumstances Example 2.2 can not be interpreted idiomatically. A literal take is necessary. Similar applications of interjections as tests of constituent structures were considered in [47] and in Chapter 4 in [48].

Example 2.2
John kicked .... uh .... the bucket. (James 1972)

Swerts *et al.* [117] discussed in detail the communicative functions and the acoustic correlates of filled pauses. Filled pauses used to have longer duration than similar vowels in normal contexts [110]. Nasal filled pauses like *um* and *un* played a different role than the unnasalized one, *uh*, regarding the acoustic features and the phrasal positions of filled pauses. Nasalized

[2] "**Oh1** indicates that the speaker has just now become aware of something covertly.... **Oh2** indicates that the speaker has now paused to make a decision or choice between alternatives... **Oh3** is similar in use to *well*... **Hey** indicates that the speaker is with thought at the time of uttering hey and desires the addressee's attention in order to place material into the shared world. **Aha** indicates that the speaker has just now covertly pieced together the logic of a situation or seen a connection previously missed... **Ah** indicates that the speaker has just now thought of something and finds that thing, or having thought of it, pleasing or significant... **Uh**,..., indicates "speaker is hesitating to try to think of the best or most accurate thing to say next, or to remember something, or is reluctant to say what is to follow."" (Schourup 1983, P. 108)

ones often appear in phrase-initial positions and are followed by longer silent pauses. Unnasalized ones, however, signal rather shorter interruptions, appearing often in non-phrase-initial positions [12].

Another communicative function that filled pauses serve is to hold the floor. This has frequently been proposed in the literature. Instead of just being silent, the speaker produces some sounds to signal to the listener that he or she has not yet finished with his/her talk and thereby keeps the floor. This hypothesis is obvious but difficult to prove. Experiments have been conducted by Lalljee & Cook [60]; [59], in an attempt to find support for the relationship between filled pauses and floor-holding. Nevertheless, they have not found direct evidence supporting this hypothesis.

Silent Pauses and Sentence Structuring In read speech, pauses possibly reflect the syntactic knowledge and realization of the speaker, whereas the duration of pauses distinguishes the sentence and clause boundaries. Goldman [34] for instance measured the duration of unfilled pauses and found out that the majority of sentences (77.9%) were divided by pauses longer than 500 msec, whereas 66.3% of interclausal pauses were shorter than 500 msec. Similar results were also obtained by Grosjean *et al.* [35]. The frequency and the length of pauses in their corpus of read sentences were related to the syntactic constituents. Interclausal pauses were less frequently used. When the pauses were used the length was smaller than that between sentences. This has been thoroughly discussed by Wightman *et al.* [125] by means of an investigation of the segmental duration in prosodic vicinities. Pauses within sentences structure the syntactic phrasing, thus providing the listener with a more understandable input [109]; [6].

Prosodic Features of Pauses That prosodic cues play a primary role in resolving the processing problems presented by disfluent speech has been supported by experiments carried out by Lickley & Bard [70]. Their results of gating experiments suggest that disfluency in spoken language could be detected before the word following the interruption has been completely recognized. The results also show that listeners were able to recognize the occurrence of disfluencies before having access to lexical information, relying primarily on prosodic and acoustic cues. These results strongly support the hypothesis that listeners clearly make use of prosodic cues, or rather, rhythmic expectancy, to detect prosodic continuity of sentences in disfluent speech.

Duez [22] investigated the perception of subjective pauses in speech and their acoustic correlates. Among the prosodic factors which affect the perception of pauses, including: vowel lengthening, fundamental frequency and intensity; the most effective prosodic cue in perceiving pauses seems to be final lengthening. The results also support the proposal of Butler [8], that the syntactic structures play a less obvious role than the intonation pattern in influencing the perception of pauses.

Shriberg & Lickley [113] made use of acoustic-prosodic information to develop a model to predict F_0 values of filled pauses. They applied the peak F_0 values of sentence-initial stressed syllables and the related baseline values. Additionally, it was also found that "filled pauses all had falling or flat F_0 patterns, at relatively low F_0 levels" which is also proven by Shriberg & Lickley [113].

$$F_0 \text{ (filled pause)} = \text{r } (F_0 \text{ peak - } F_0 \text{ baseline}) + F_0 \text{ baseline}$$

The prosody-particle pairs suggested by Gibbon & Sassen [32] are concerned with the phatic functions of discourse particles, especially interjections (filled pauses). Types of stylized intonation were integrated into lexical entries formalized in HPSG conventions, focusing on the attributes dealing with communication channel (CC)-oriented considerations. In their suggestion, the prosody discourse pairs were explicitly integrated into a formal lexicon.

2.1.4 Speech Errors

The investigation of speech errors (slips of the tongue) has a long history. The first large collection of speech errors of German was done in 1895 by Meringer & Mayer [80]. This method is still widely adopted, for the reason that one obtains more cues about human speech processing through their anomalous realizations than through "normal speech". As there are a large variety of classifications suggested for speech errors, a classification proposed by the author should illustrate the features inherited in a classification of speech errors.

Phonological errors:
bleiden (blauen+beiden) [3], F́ortsetzung (Voráussetzung) [4]
Morpho-syntactic errors:
steht (stehen) [5], die zweiten (die zweite) [6]
Lexical errors:
die orangene Mutter (die gelbe Mutter) [7]
Syntactic errors:
das (der Bug)[8], vor (an den gelben Würfel)[9]

Phonological errors are errors whose syllabic structure is similar to that of the intended lexical entry but with either different prosodic features (stress or quantity), or false segmental realizations. Stress errors can be made while building derivations, e.g. *artikulatórisch - Artikulatión*, in which the word stress is shifted by word formation rules. Phonological errors are not necessarily legal lexical units in a given lexicon.[10].

[3] blauen = blue, beiden = both
[4] Fortsetzung = continuation, Voraussetzung = prerequisite
[5] steht = stand (3rd, singular), stehen = stand (plural)
[6] die zweiten = the second (plural), die zweite = the second (feminine, singular)
[7] die orangene Mutter = the orange nut, die gelbe Mutter = the yellow nut
[8] das der Bug = the (neuter) der (masculine) nose
[9] vor an den gelben Würfel = in front of onto the yellow cube
[10] The occurrence of phonological errors can be in the form of addition, anticipation, blend, exchange, haplology, malapropism, omission, perseveration and substitution, de-

Example 2.3
Und mhm jetzt die bleiden blauen Schrauben. (Sagerer et al. 1994)
(And uhm now the bloth[11] blue bolts.)

Example 2.4
Aber daß die ähm, daß die blei beiden ähm vorne diese Propellerteile da ähm so diagonal sind. (Sagerer et al. 1994)
(But that the ehm, that the blo both ehm in front these propeller parts over there are ehm so diagonal.)

In Example 2.3 *bleiden* is not a lexical entry in German. It is an anticipation error composed of "beiden" (both) and "blauen" (blue). "Blei" (lead) is a lexical entry in Example 2.4 contrary to "bleiden", but according to the usage in this context, "blei" can not be the word "Blei" (lead) but a speech error where an /l/ is inserted in the word "beiden" and the speaker has interrupted the word before the word has been completed after detecting the error.

Morpho-syntactic errors are segment strings which deviate from the correct morphological forms in their syntactic context. Morpho-syntactic errors are usually of the lexicon form. The listener is usually not supposed to be able to detect this kind of error, but the speaker, by contrast, generally detects and corrects the errors during the self-monitoring of his speech [68]. Morpho-syntactic errors are syntactic errors made at the morphological level: false declension (Example 2.6), false ending and false conjugation in subject-verb-congruence (Example 2.5).

Example 2.5
Stehen steht dieses ähm Fünf diese Schiene parallel zu der großen oder? (Sagerer et al. 1994)
(Do does this ähm five this track stand parallel to the big one or what?)

Example 2.6
Und dann nimmst Du die zweiten, die zweite von diesen Sieben-

tailed examples see [16].
[11]It is a hypothetical translation.

erlöcherndingern da. (Sagerer et al. 1994)
(And then you take the second, the second one out of those with seven holes over there.)

Lexical errors are errors that occur at the lexical level such as word substitutions. Lexical errors convey false or inappropriate meaning which does not correspond to the semantic content of the intended word. Word substitutions are usually content words such as nouns, adjectives and adverbs (see Example 2.7).

Example 2.7
Und unten runten ist halt die gelbe Mutter äh die orange Mutter. (Sagerer et al. 1994)
(And beneath that is the yellow nut eh the orange nut.)

Syntactic errors: incorrect articles (Example 2.8) and incorrect cases are more often to be corrected right after the errors are detected by the speaker than the morpho-syntactic errors. Syntactic errors include grammatical words whose syntactic context is false. Syntactic errors do not have to do with derivation or composition of words. In addition, false word order is also a kind of syntactic error.

Example 2.8
Das wird jetzt das der Bug. (Sagerer et al. 1994)
(That will now be the[neuter] the[masculine] nose/front.)

2.1.5 Speech Repairs

In dealing with speech repairs researchers have been able to obtain more linguistic indications than merely investigating speech errors, as speech repairs logically contain speech errors. Speech repairs only occur when some words are incorrectly or inappropriately produced. Pauses, filled or silent ones, can often be observed within speech repairs which have complex internal structures. Speech repairs were examined from various perspectives in the 70s and 80s [104]; [102]; [109]; [66]; [67]. In particular, the classification of speech repairs must first be made available to carry out specific investigations. And how speech repairs are classified depends on the goals of the individual research work. Thus, this section focuses on the classification of types of speech repairs which have been suggested in the literature, and on some important research work on speech repairs such as speech monitoring and modelling.

Speech Repairs and Monitoring Levelt [66] suggested three phases in which self-repairs typically proceed:
1) the monitoring of one's own speech and the interruption of the speech flow when a problem is detected,
2) hesitation, pausing, but especially the use of so-called editing terms and
3) making the repair proper.
In his explanation, Levelt mentioned disfluent elements which are used, or rather, appear in spoken language: *trouble-producing*, *trouble-detecting* and *trouble-correcting*. The occurrence of the troublesome strings provides insights into the monitoring processing of speakers, as mentioned in MacKay [77], Fromkin [29], Dell & Reich [20] and Levelt [68]. Moreover, it gives useful insights into the production of speech. The position where speakers re-initiate their troublesome speech strings [66] may be cognitively relevant to language processing, if significant regularity can be found.

Classification of Repairs Repairs can be classified by means of linguistic levels such as lexical or syntactic ones, as done by Shimanoff & Brunak [109] and Levelt [66]. How many categories one needs for a classification proposal is dependent on the specific purpose of investigations. Therefore, the number of criteria does not necessarily mean the extent of the classification can cover. Another possibility is to classify them by means of structural descriptions, as suggested by Blackmer & Mitton [5]. This section introduces these three different approaches mentioned above.

Shimanoff/Brunak (1977) Shimanoff & Brunak [109] classified repairs according to which level they occurred at during discourse: phonological, morphological, lexical, syntactic and/or discourse, as defined in Table 2.2:

Table 2.2: Shimanoff/Brunak (1977)

repair types	**realizations**
phonological repairs	a change in pronunciation or the substitution of sounds
morphological repairs	alterations in minimal units of meaning
lexical repairs	substitution of one word for another
syntactical repairs	word order or word agreement
discourse repairs	involvement of trouble source across two or more clauses.

Levelt (1983) Levelt [66] differentiated four kinds of repairs, as shown in Table 2.3

Table 2.3: Levelt (1983)

repair type	**occurrence**	
(error)	E-repairs:	42%
(covert)	C-repairs:	25%
(appropriateness)	A-repairs:	30%
(different)	D-repairs:	1%
(rest)	R-repairs:	2.5%

E-repairs are repairs which correct errors including repairs of syntactic errors (ES-repairs 2%), repairs of lexical errors (EL-repairs 38%) and repairs of phonetic errors (EF-repairs 1%). Covert repairs (C-repairs) in Levelt's classification can be interruption plus editing terms and repeats of some lexical items. A-repairs, appropriateness repairs, are some further semantic specification modifying the current messages which are inappropriate in the speaker's opinion. When the problem terms are replaced by different ones, these different ones are called D-repairs. D-repairs are in fact restarts. R-repairs are the rest of repairs which can not be identified as any of the above categories.

Blackmer/Mitton (1991) Blackmer & Mitton [5] differentiated covert repairs from overt repairs. Their classification is based on the structural properties of speech repairs, rather than on linguistic levels, as used by Shimanoff & Brunak and Levelt. As shown in Table 2.4, covert repairs include editing terms and repetitions, being located in within-utterance or between-utterance positions. Among the conceptually based overt repairs, appropriateness repairs are differentiated from different repairs. Likewise, a distinction is made between the conceptually based and the production based overt repairs, although it is difficult to draw a clear line between these two groups.

Modelling Speech Repairs Another aspect of speech repairs is the modelling of their production. Connectionist modelling done by Schade [99] and Schade & Laubenstein [100] showed that results of the modelling corresponded to a great extent to the data found in real speech. Schade and Schade & Laubenstein used a special monitoring component to detect inconsistencies between the current production and the intended utterance which subsequently de-activates the flow of production. The monitor then initiates a repair on the level where the de-activated node was. The behaviour of the model with respect to the initiation position of repairs is described in [100] as follows:

Case 1. If the utterance already produced is continued beyond the item of error, the whole item at a hierarchically higher level will be repaired. The following is an example: *Take the green block eh the red block* and *Take the*

Table 2.4: Blackmer/Mitton (1991)

Covert Repairs
- within-utterance covert repair
within-utterance editing term
repetition without editing term
repetition with editing term
- between-utterance covert repair
Overt Repairs
(with or without editing terms or repetitions)
- conceptually based overt repair
appropriateness repair
appropriateness insert
appropriateness replacement
different repair
- production based overt repair
- unclassifiable overt repair

green block eh take the red block.
Case 2. If the utterance is interrupted directly after the production of the erroneous item, repairs such as *Take the green eh red block* and *Take the green eh the red block* are more likely.
Case 3. If the utterance is interrupted during the production of the erroneous item, a repair such as *Take the gr/ eh red block* is more likely.

2.2 Structural Aspects

Because speech repairs are more complex than the other disfluencies and can therefore provide more information about the internal structure of disfluencies, the discussion of the structural aspects of disfluencies will focus mainly on speech repairs. In the three different classification types of speech repairs introduced above, the internal structure of speech repairs has not explicitly been taken into consideration. Most classifications of speech repairs stress the importance of linguistic level on which speech repairs take place. This section will deal with research results related to the internal structure of speech repairs. The internal segmentation of speech repairs can be considered a sequence of three elements: reparandum, edit signal and repair, as conventionally suggested in the literature [66]; [82]; [110]; [114]. In Figure 2.1, the *reparandum* is a string of words or phrases which are in the speaker's opinion errors or inappropriate and therefore should be corrected. Sometimes in spoken language or conversation they are not corrected at all. The reparandum is very often an incomplete and interrupted word or phrase. The items which correct the reparandum are *repairs.* Repairs can appear as words or phrases, depending on the context of how the reparandum should be corrected. *Edit signal* occurs between the reparandum and the repair and can be silent/filled pauses or lexically marked sequences. Speakers frequently interrupt their speech immediately after the reparandum (phrase or word); sometimes they even interrupt the speech within the reparandum itself. Whether the phrase has already been completed doesn't seem to play an important role. This has been called **Main Interruption Rule**, "Stop the flow of speech immediately upon detecting the occurrence of repair", by Levelt [66].

Regarding the classical segmentation of speech repairs *(reparandum, edit signal, repair)*, there are two approaches dealing with the structural description of disfluencies: declarative and procedural. The approaches suggested by Levelt [66], Nakatani & Hirschberg [83], Shriberg *et al.* [4], Shriberg [110]

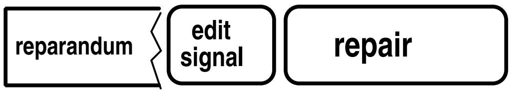

Figure 2.1: Internal Segmentation of Speech Repairs

and Clark [13] give declarative descriptions of the occurrence of disfluencies. Hindle [42], used a procedural mechanism to detect repairs in the context of building a deterministic parser. Several systems for describing the internal structure of speech repairs will be introduced in this section. Description methods suggested by Levelt [66], Shriberg [110] and Clark [13] focus more on the flat surface structure of speech repairs, whereas the system used by Bear *et al.* [4] is intended to be applied to label speech repairs in spoken language. This sort of labelling system is often combined with a pattern matching method to detect and correct speech repairs, as done in [4] and [39].

2.2.1 Segmentation Systems and Comparison

The location of where the reparandum and the repair begin and end vary from definition to definition. As shown in Figure 2.2, the stretch being replaced is the reparandum and the stretch replacing the reparandum is called the correction. Levelt used "delay" to indicate the number of syllables after the reparandum to the beginning of the interruption and "retracing" to mean the number of syllables in the original utterance which are repeated after the editing phase. In this case, the value of delay is three and the value of retracing is one.

In contrast with Levelt, Shriberg's system [110] includes the retraced elements in the domain of the reparandum and the repair. The location of the interruption is emphasized in her segmentation as well as in Levelt's model. As shown in Figure 2.3, Shriberg introduced a rather simple segmentation criterion, concentrating on the elements after and before the edit terms.

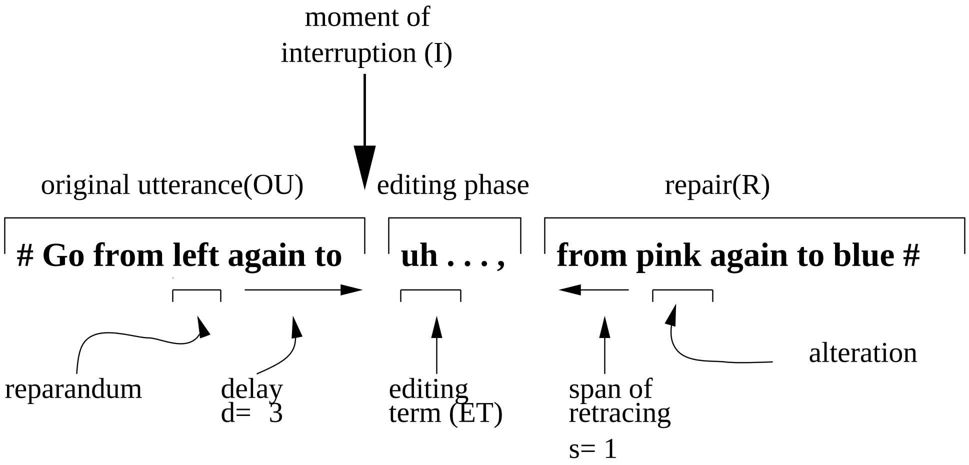

Figure 2.2: Levelt (1983)

The disruption schema suggested by Clark [13] stresses the relevant sites within speech repairs, as illustrated in Figure 2.4. Clark viewed disfluencies mainly from the perspective of the suspension and resumption sites. Where the reparandum begins and where the repair ends are not of main interest in his disruption schema. Although he uses different terms for the intervals defined in his schema, no great distinction is made with respect to the other models: hiatus for editing terms in [66] and for interregnum in [110]; suspension point for moment of interruption etc.. The weakness in his suggestion is that the beginning of the original delivery and the end of the resumed delivery are not clearly pointed and defined. For instance, the original delivery is defined as the smooth delivery before the disruption, which can be retraced to a further back position prior to the interruption where the string has been smoothly uttered. That is, the boundaries of the intervals, the very beginning and the very end, which are nevertheless relevant to the production of speech repairs are not clearly defined.

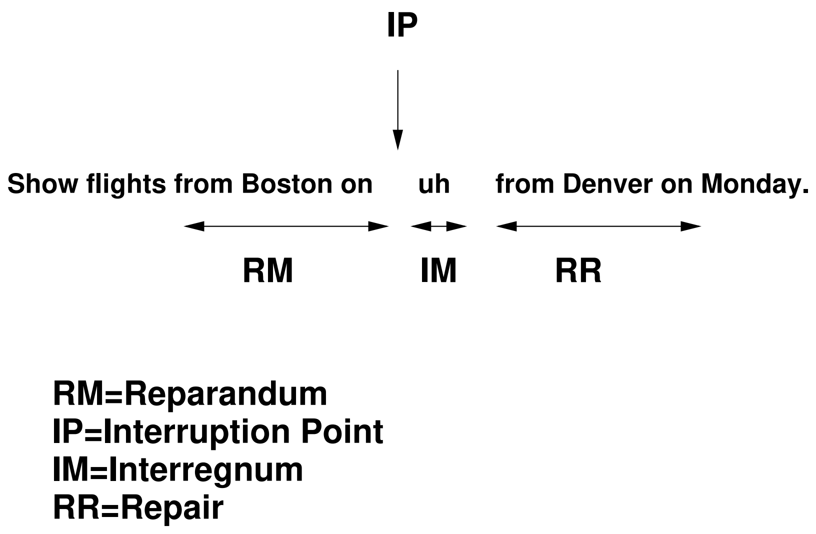

Figure 2.3: Shriberg (1994)

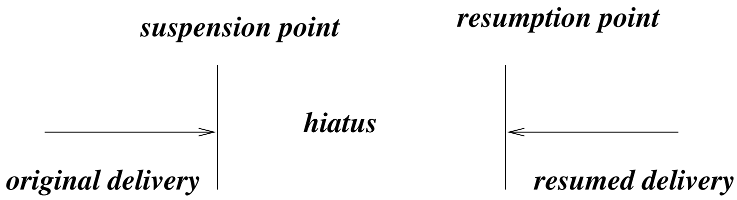

Figure 2.4: Clark (1997)

Comparison

The property shared by the three approaches introduced above is that they all stress the importance of the location of interruption and take it as a marking site for the offset of the erroneous strings and possible onset of the repair strings, as illustrated in Figure 2.5. The descriptions suggested by Levelt and Shriberg have more in common than the one proposed by Clark. In Levelt and Shriberg's segmentation, three intervals can be clearly distinguished: reparandum, editing phase and correction. However, the length of these three intervals is differently defined. Levelt focuses more on the word which is *really* corrected, whereas Shriberg takes the string and retraces it back to the word which is repeated after the interruption.

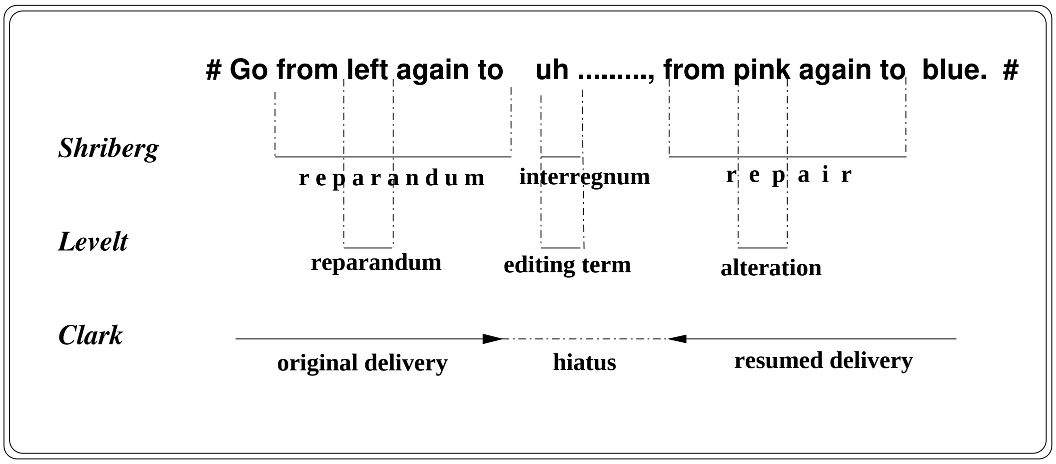

Figure 2.5: Comparison of Levelt, Shriberg and Clarks Segmentations

Comparing these three approaches, we can conclude that there are locations which can be relevant to the internal structure of speech repairs:
1) the location of interruption,
2) the problem word/s before the interruption,
3) the corrected word/s after the interruption and
4) the beginning word after the interruption.

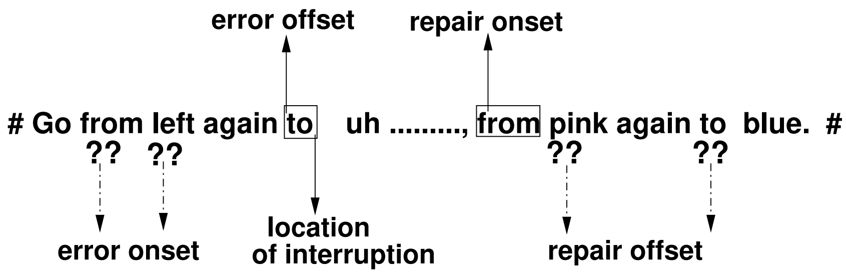

Figure 2.6: Terms Used/to be Investigated in this Thesis

In this thesis, as shown in Figure 2.6, we use the term *location of interruption* for the position where the flow of speech is interrupted. *Error offset* stands for the last word before the interruption, whereas *repair onset* represents the first word after the interruption. This has been consistent throughout the three approaches mentioned above. But where the error begins (*error onset*) and where the repair ends (*repair offset*) vary to a great extent among those definitions above. Error onset can be defined as the same/corresponding word as the repair onset or alternatively as the problem word itself, i.e. the "reparandum" in Levelt's definition. Similarly, repair offset can be used for the same/corresponding word as the error offset or the *really repaired word*. This is called the "alteration" in Levelt's definition. Therefore, before the syntactic analysis in Chapter 3 can be carried out, the boundaries of speech disfluencies, namely error onset and repair offset, must be first clearly defined. Exactly these two positions will be investigated in the next chapter to see whether one can use certain boundary specification to describe these two locations consistently.

2.2.2 Labelling Systems

Bear *et al.* [4] too adopted a kind of declarative approach to describing speech repairs, as shown in Table 2.5. It was not intended to develop a description system for speech repairs, but to attempt to catch certain types of occurrence of speech repairs and to annotate them. They divided speech repairs according to the length and the surface structure of repairs. What is meant by length is the number of words which are troublesome and should be corrected later. Operations which can be observed in the surface structure are repetitions, insertions and replacements.

A similar labelling system was used by Heeman & Allen [39] to apply their statistical language model to detect and correct speech repairs. Like Bear *et al.*, they were also concerned with word fragments, word repetitions and word replacements. Longer word strings were not taken into account. Only one-word and two-word repetitions (**m-m, mm-mm**) and adjacent replacements (**rr**) were considered in their annotation system.

Table 2.5: Bear/Dowding/Shriberg (1992)

Type	Pattern
	Length 1 Repairs
Fragments	M_1-, R_1-, X-
Repeats	$M_1\|M_1$
Insertions	$M_1\|X_1....X_iM_1$
Replacement	$R_1\|R_1$
Other	$X\|X$
	Length 2 Repairs
Repeats	$M_1M_2\|M_1M_2$
Replace 2nd	$M_1R_1\|M_1R_1$
Insertions	$M_1M_2\|M_1X_1....X_iM_2$
Replace 1st	$R_1M_1\|R_1M_1$
Other	...|...

2.2.3 A Detection Mechanism

Labov [57] proposed that a set of simple rules of interpretation should account for certain types of disfluencies in spoken language. These rules should model the *regularity* of the "irregular" everyday speech. Following Labov's argument, Hindle [42] made use of the existence of edit signals as an indication of the end of the reparandum and the beginning of the repair, based on the *edit signal hypothesis*. Edit signal hypothesis assumes that a certain edit signal always marks the location where self-corrections take place [57]; [42]. The parser has a push down stack of incomplete nodes and a buffer of complete constituents, into which the grammar rules can look through a window of three constituents. The deterministic parser in which a mechanism is provided to detect and correct speech repairs is able to process spoken language data. An important presumption regarding the parser is the existence of edit signals. Based on the location of edit signals, the parser applies a copy editor mechanism for detecting a repair structure. A repair structure is detected when on both sides of edit signals the same words or words of the same category occur. After detecting a repair structure, the parser expunges the word strings before the edit signals. As shown below, the *surface copy editor* and the *category copy editor* compare the surface structure and the categories of words on both sides of edit signals, whereas an edit signal is marked by –. If the occurrence or the category of both sides are identical, it is a repair structure.

surface copy editor

Well *if they'd–* if they'd had a knife I *wou–* I wouldn't be here today.
If they– if they could do it.

category copy editor

I was just *that–* the kind of guy that didn't *have–*
like to have people worrying.

The parser has a third copy editor: the *stack copy editor*. If an edit signal precedes the first constituent in the window, the stack copy editor looks through the stack and expunges constituents which have the same category as that of the first constituent in the window. The stack copy editor can deal with utterances like the following:

stack copy editor

I think that *you get–* it's more strict in Catholic schools.

Restarts can also be recognized by the parser. They can be identified by means of the lexically marked strings such as *well* or *ok*.

lexically triggered restarts *well, ok, see etc.*

That's the way if– well everybody was so stoned, anyway.
But when I was young I went in– oh I was nineteen years old.

Another possibility to detect speech repairs is to use the pattern matching approach. Bear *et al.* [4] and Heeman & Allen [39] both applied this method to detect speech repairs, whereas Heeman/Allen [39] applied repair pattern rules combining them with a statistical model filter to process speech repairs. Heeman & Allen distinguished three types of speech repairs regarding their structure: fresh starts, modification repairs and abridged repairs. Fresh starts are defined in the same way as restarts, as mentioned earlier. A modification repair modifies what was said previously, whereas an abridged repair is only composed of a fragment or editing term. These three types of speech repairs were annotated in a similar system to the one used by Bear *et al.* [4], but with additional part of speech (POS) information. Applying prescribed repair pattern rules, the pattern matching method was used to detect the repairs.

In the context of detection and correction of speech repairs, the work done by Langer [62] should also be mentioned. He suggested a set of normalization rules to deal with ungrammatical sequences in spoken language. He was concerned with *explicit repairs* and *ungrammatical repetitions*. His parser scans the word strings, and if a repair indicator (similar to the edit signal discussed in the present context) is found, the word preceding the indicator is then removed. The process runs until the parser accepts the word strings.

2.2.4 Discussion

In the declarative and procedural approaches outlined in this section, it is apparent that there are three important intervals within speech repairs which shape the main structure of repairs: reparandum, edit signal and repair. In the approaches suggested by Hindle and Langer, it is presumed that edit signals or repair indicators must first be identified. Only based on this presumption, can the detecting rules be applied. In the psycholinguistic research fields related to the existence of the edit signal, Lickley *et al.* [72] could not find experimental support for the function of the edit signal to reveal disfluencies (repetitions and restarts) where they are claimed to occur. Using gating experiments, the majority of the subjects could detect the occurrence of disfluency by the offset of the first word after the disfluency, whereas only a few subjects were reported to be able to detect the oncoming continuation using the phonetically identifiable edit signals. However, for computational linguistic approaches, the detection results have proved to be good if the location and the realization form of edit signals are used.

Another issue regarding the structure of speech repairs is the definition domain of individual intervals. The length and terms of the three intervals mentioned above, *reparandum*, *edit signal* and *repair*, are differently defined, according to the goals of the approaches. But it is clearly observable that there exist certain locations which are particularly relevant to the internal construction of speech repairs. These are the interruption site, the offset of the reparandum and the onset of the repair. These three locations play an important role not only in description systems, but also in detection mechanisms. In Hindle's approach, the constituent preceded by an edit signal is checked in terms of the type of constituent following the edit signal, because edit signals mark the possibility that the preceding constituents may be inappropriate. It is the same in Langer's case; the element preceding the repair indicator is removed if the parser cannot parse the strings.

2.3 Prosodic Aspects

Concerning the prosodic properties of disfluencies, two research fields have frequently been investigated: the prosodic realization of the production of disfluencies and the detection of disfluencies by humans or machines using prosodic information. Prosodic characterizations of the production of disfluencies help explore human language processing. One can obtain indications of the relationship between prosody and other linguistic devices, for instance, between prosody and lexicon, prosody and semantics as well as prosody and syntax. Examining the production of disfluencies, especially their prosodic properties, does not only have applications in psycholinguistic research fields as mentioned, but also in computational linguistic approaches, namely natural language processing (NLP). Acoustic-prosodic cues also provide important and efficient information about the occurrence of disfluencies. In automatic speech recognition, making use of prosodic information and the structural properties mentioned in the previous section, disfluencies can be detected more efficiently. How human beings detect the occurrence of disfluency is also an interesting issue which can be examined by means of perception experiments. Experiments carried out by Lickley & Bard [70]; [74] showed that subjects in the gating experiments could detect the disfluent speech stretches within the first word after the interruption, without knowing what the first word was. Prosodic cues seem to support the detection of disfluencies more than lexical cues.

Below, recent studies and analyses of disfluencies focusing on speech production are to be briefly summarized. By investigating the prosodic realization of disfluencies, one can obtain new insight into the constitution of the lexicon with respect to the role played by prosody. Next, research results concerning the relationship between prosody and semantic contrast will be introduced. Examining the prosodic production of speech repairs, Levelt & Cutler [69] and Levelt [67] have found a clear correlation between semantic contrast and the prosodic form of speech repairs and editing terms. Afterwards, investigations dealing with the reset hypothesis are summarized. Approval of the reset hypothesis has applications both in psycholinguistic and computational linguistic fields. Having at least the same effect in improving automatic speech recognition results, temporal features have also been investigated in detail, especially those of pauses. Results concerning temporal features of repeats and speech repairs are discussed. Recently, in NLP acoustic-prosodic cues have proved to be useful in identifying disfluencies. Approaches and investigations related to acoustic-prosodic cues of disfluencies are subsequently reviewed.

2.3.1 Speech Production

By means of systematizing the occurrence of disfluencies, one can find out what syntactic or prosodic forms, and what positions within utterances are preferred by the speakers in the production of disfluencies. This section is primarily concerned with research results investigating lexical-prosodic cues and speech processing units through speech error data.

Prosody and Lexicon Dell & Reich [20] introduced the concept "lexical address" which specifies where the stem of each word is to be found in the listing in a lexicon. So as exemplified by Dell & Reich, the lexical addresses of "present" and "pressure" may be very close or similar to each other. The stage at which an error is made can be because of the following interferences: in the mapping from meaning to lexical address or in the mapping from lexical address to word, but not both. This consideration gives some further points in examining how lexical selection is to be connected with speech error data.

Van Wijk & Kempen [126] suggested that there are two mechanisms for computing the shape of self-repairs: reformulation and lemma substitution. Reformulations are addition repairs, whereas lemma substitutions are represented by substitution and deletion repairs. The prosodic qualities of lexical entries become important in the case of lemma substitution. The answer to the question *how the local prosodic features are changed, when repairs occur in the lexical domain* gives lexical-prosodic cues to how prosodic features should be integrated into a lexicon. In [68], Levelt has proposed a well-formedness rule for repairs[12]. According to the results of the analysed

[12] Well-Formedness Rule for repairs
An original utterance plus repair OR is well formed if and only if there is a string C such that the string $OCorR$ is well formed, where C is a completion of the constituent directly dominating the last element of O (or is to be deleted if that last element is itself a connective such as *or* or *and*). (Levelt 1989)

corpus in [126], Levelt's well-formedness rule, which connects self-repairs to co-ordinate structures, is shown to apply only to reformulation. The reason is that the important linguistic unit in reformulations is the major syntactic constituent, whereas in lemma substitutions it is prosodic units (the phonological phrase) which play a significant role.

Unit of Speech Processing It is often argued that most of the speech production models assume that the word or the lexical item is the unit of language processing. But in [108], the speech error data analysed by Shattuck-Hufnagel strongly suggested that words and morphemes are not necessarily the most suitable processing units, but rather smaller units such as syllables or part of syllables would be more appropriate. Other evidence offered for example by Fromkin [29] also provided support for the fact that the phonemic segments can be units of speech processing as well. Fromkin [29] suggested that the largest percentage of speech errors show substitution, transposition, omission and addition of segments of the size of a phone. They occur within words or across word boundaries. In [108], it was found that where the intended words included consonant clusters, only one consonant of the cluster was involved in the associated speech errors. This result supports the notion that individual segments are units of speech production. When vowels or syllables or parts of syllables or whole words are substituted or transposed, there is no change in the stress pattern or the contour of the sentence. It should be noted that, whether a local disfluency can be found was not investigated.

Besides the proposals of words as processing units and segments as processing units, Shattuck-Hufnagel [108] suggested that syllables can be a stable processing unit for speech production. That syllables play a significant role in speech production planning is also suggested by Dell [19] and MacKay [77]. Because of the interest in syllables nowadays in determining the allophonic shape of a segment and of insights into autosegmental phonology and metrical phonology, it is justified to re-examine lexical and phonological errors to see whether syllabic sub-constituents are appropriate to serve the role of a basic unit. According to the examination of the speech error data, Shattuck-Hufnagel only proved the fact that word-onset consonants function as a special processing class, but still could not motivate a representation of the full syllabic substructure of the utterance during planning.

2.3.2 Prosody and Semantic Contrast

The concept *prosodic markedness* can be referred to various interpretation possibilities and research interests when prosodic features of disfluencies in spoken language are considered. This section summarizes some influential results with respect to the prosodic markedness of disfluencies and related problems. Prosodic markedness can be used in various forms. Levelt & Cutler [69] investigate the prosodic markedness of repairs, elicited from pattern description monologues. The two authors separately marked the prosodic changes by auditory judgement and afterwards compared their results to examine the extent of agreement. The cases agreed to by the two authors were then used to evaluate the examination.

Prosodic markedness defined in the study includes differences in loudness, contour changes and relative duration. Evaluating the results, Levelt & Cutler [69] found that the erroneous utterances are more frequently prosodically marked than the inappropriate ones. The phenomenon *prosodic marking of speech repairs* has been discussed by Levelt & Cutler [69] on the relevance of semantic contents on prosody. The prosodic marking is supposed to serve the function of accenting the troublesome terms the speaker has made and to help the listener to be aware of the situation and to solve his continuation problems when the speaker has interrupted his speech. “A correction is marked when the prosody of repair item and trouble item differ.” (defined by Levelt & Cutler [69]). Two main determinants should have an influence on the usage of the prosodic marking of repairs: syntactic and semantic. When the interruption is delayed, the number of words or phrases in the original utterances repeated decides how accenting the speaker wants the listener to perceive. The semantic differentiation between *the reparandum* and *the repair* functions the same, to emphasize the peculiarity of repairs prosodically. Furthermore, they tried to convey the different degree of prosodic marking of erroneous items and inappropriate items. In their perceptual corpus analysis, 53% of speech errors were prosodically marked, whereas only 19% of inappropriate items were prosodically marked. They came to the conjecture that the semantic contrasts of the reparandum and the repair play a certain role in the prosodic markedness. Prosodic marking is more frequent if the contrast of contextual alternatives to the rejected item is small.

Levelt [67] examined the correlation between editing terms and prosodic markedness. 55% among the "contrast establishing" editing terms *nee* (no), *of* (rather), *sorry* (sorry) and 32% among the "neutral" editing term *eh* (er) and the non-contrasting *dus* (therefore) were prosodically marked in the corpus. This result also supports the existence of the relationship between semantic content and prosodic means. The contrast establishing editing terms are used to emphasize the semantic difference, whereas the non-contrasting editing terms carry a less contrastive function. Editing terms and prosodic marking seem to arise from a semantic level of representation, but this does not exclude the existence of other determinants.

2.3.3 Reset Hypothesis

Levelt [67] suggested that one would not expect the repair to be an acoustic continuation of the interrupted original utterance, though grammatical contiguity would be possible. It is well known that the fine details of acoustic shape of speech are not represented in working memory; the level of coding is phonemic or semantic rather than phonetic. It is predicted by Levelt that "acoustically the new utterance should fit seamlessly into the original utterance, even for longer delays (or at least there is no known limit on the persistence of an interrupted speech programme)" [67]. This prediction is similar to the content of the reset hypothesis. The reset hypothesis predicts F_0 will be reset after the interruption, to a higher value than expected, following the baseline of a fluent utterance. Usually, utterances obey the baseline declination. When it is repaired during an utterance, F_0 is reset to the value of a certain previous position. This position is very likely to be the corresponding position of the repair initiation.

Shriberg's [111] results illustrated that only few prospective repeats (cf. Section 2.1.2) showed the effect of the F_0 reset, whereas retrospective repeats (cf. Section 2.1.2) much more often showed. This result proved the function and possible existence of F_0 reset within disfluencies.

2.3.4 Temporal Features

Temporal characterizations such as duration or speech tempo are particulary important, when spoken language is concerned [18]. Temporal features of disfluencies can be examined from a variety of aspects. For instance, duration of the three main intervals within disfluencies can be measured, namely the reparandum, the edit signal and the repair, as shown in Figure 2.1, or the duration of words within disfluencies which are related to each other with respect to their position in the reparandum and the repair. Another possibility is to examine the relationship between speech tempo and the rate of disfluencies. However, most of the examinations performed were related to the first case. For instance, Blackmer & Mitton [5] not only gave a clear classification for repairs from the perspective on the structure of repairs, but also analysed disfluencies in terms of their temporal distribution.

Error-to-cut-off (the interval between the beginning and the end of the error), cut-off-to-repair (the interval between the end of the error and the beginning of the repair) and error-to-repair (the interval between the beginning of the error and the beginning of the repair) times were measured by Blackmer & Mitton, who attempted to find some evidence for or against the flow-through monitoring by Levelt [68] and the hold-up monitoring by

Laver [63]. In most of the language production models, it is supposed to be observable that speakers need time to mentally process and organize their speech when errors have been made. The results achieved by Blackmer & Mitton [5] show that 0 ms cut-off-to-repair times appeared in 12,4% of the overall sample and in 19,2% of the overt repairs. The fact that one-fifth of the overt repairs were realized without being accompanied by *processing time* contradicts the idea that speakers need to re-process and re-build their erroneous sequences. The results support the concepts of incremental planning and buffering by the articulator, whereas the need for silent glottalization planning time by the articulator prior to initialization of a segment is not supported. However, only to a certain extent, does this support the notion that people can process their speech incrementally and initiate multiple tracks for cognitive processing. This is because it is not clear whether the time of articulating the error itself is used for retrace processing.

Shriberg [111] made use of the durational intervals of repeated words to find evidence supporting the hypothesis suggested by Hieke [41]. In examining direct repeats, the durational lengths of the words contained in the repeats provide an indication of the classification of types of word repeats: prospective and retrospective [41]. Shriberg [111] supplies empirical evidence supporting the difference between these two kinds of repetition types [41], namely, both the difference of the location of pauses and the distribution of duration of the two types of repeats.

2.3.5 Acoustic-Prosodic Cues in NLP

In addition to the conventional structural aspects of speech repairs, Bear *et al.* [4] used acoustic techniques to develop a multiple knowledge model for detection and correction of speech repairs in spoken language. They used a pattern matching technique for the most frequent occurrences such as fragments, repeats, insertions and replacement. They worked with transcript material, which is a text-first model in contrast to the speech-first model suggested by

Nakatani & Hirschberg [82], and found matched patterns. Different analyses were performed with additional syntactic and semantic knowledge and better results were obtained. They also discovered that prosodic-acoustic cues can be useful in this context, for example, heavy glottalization at the end of a fragment which appears different from normal fluent speech. Bear *et al.* [4] therefore proposed a multiple knowledge model of automatic detection and correction of speech repairs, but the few patterns which they found in their corpus can not cover all disfluencies. For instance, interjections which occur within sentences or within sentence boundaries were not included.

Daly-Kelly [17] used phrasal and temporal information about the pause intervals, defined by herself, combined with the position where vowels occur in hesitation and filled pauses. She achieved good prediction rates in human/machine dialogues by using the following information: vowel lengths are shown to be much longer in pause intervals than in normal phrases and hesitated vowels are more likely to occur in the initial or final syllable, whereas filled pause vowels usually occur in the utterance-initial position.

Nakatani & Hirschberg [82] used acoustic-prosodic cues of speech repairs based on the edit signal hypothesis (cf. Section 2.2.3) applied by Hindle [42] to detect and correct repairs. The Repair Interval Model (RIM) was based on Levelt's classification [66] which differentiated three intervals, namely, a full repair-reparandum Interval (RI), a disfluency Interval (DI) and a repair interval (RI). The interruption site (IS) too was taken into consideration. To detect the interruption site (IS), they made use of the following acoustic-prosodic cues:

a. duration of pause between w_i and w_j;
b. occurrence of a word fragment within $< w_i, w_j >$;
c. occurrence of a filled pause in $< w_i, w_j >$;
d. amplitude peak within w_i;
e. amplitude of w_i relative to w_{i-1} and to w_j;
f. F_0 of w_i;

g. F_0 of w_i relative to w_{i-1} and to w_j;
h. whether or not w_i was accented, deaccented, or deaccented and cliticized
where
w_i: lexical item to the left of potential IS
w_j: lexical item to the right of potential IS.

As shown in their corpus analysis and implementation, Nakatani & Hirschberg [82] established that acoustic cues of phrasal boundary are useful and efficient in detecting self-repairs offsets. On the contrary, edit signal hypothesis could not be supported by the corpus analysis, since only 9.8% repairs were accompanied by pause fillers or cue words.

2.3.6 Discussion

Perception experiments as well as acoustic-prosodic investigations have provided a variety of clues to psycholinguistic and computational linguistic applications. To systematize the research results introduced and summarized in this chapter, three prosodic properties have mainly been investigated:

1. F_0 values within disfluencies,
2. temporal features on various levels and
3. intonational patterns within and around disfluencies.

F_0 values provide information about the distribution of pitch and therefore help detect the occurrence of disfluencies in automatic speech recognition. F_0 values of speech repairs [4], speech repetitions [111] and pauses [87] for instance, have been used to improve the recognition results of speech recognizers. Durational properties of disfluencies, in particular pauses, are useful in both locating disfluencies and in structuring utterances.

By investigating the intervals of errror-to-cut-off, cut-off-to-repair and error-to-repair (definition cf. Section 2.3.4) in the case of speech repairs, Blackmer & Mitton [5] have obtained evidence for an incremental speech processing model. Shriberg [111] explored the durational relationship of pauses in two different groups of repeats, prospective and retrospective repeats, and has found a distinctive difference in their temporal values.

Intonational patterns can make the semantic contrast within disfluencies clearer for listeners and can also be used to segment utterances. Levelt & Cutler [69] and Levelt [67] found a clear relationship between intonational patterns and semantic contrast in the case of speech repairs.

Accordingly, these three features will be further investigated in the acoustic-prosodic analysis later on in this thesis, as well as exact acoustic measurements and prosodic properties found in Section 4.1.

2.4 Conclusion

Two questions to be asked in this thesis mentioned in Chapter 1 are:

1) How are speech disfluencies produced from the linguistic point of view, focusing especially on the syntactic and prosodic features ?
2) Do regular internal relations exist within speech disfluencies and if so, what do they look like ?

In order to obtain detailed syntactic and acoustic-prosodic cues for describing the production of speech disfluencies and therefore for detecting them, empirical analyses have to be carried out. Much attention has been given to the surface structure of speech disfluencies in most of the research approaches mentioned above. However, the focus has been more on the linearly organized word sequences, instead of on more general syntactic properties such as phrasal structure and boundaries. Therefore, the following syntactic analysis aims to investigate speech disfluencies, not merely as word sequences containing disfluency, but as integrated units with a specific internal structure. Phrasal boundaries and categories related to the disfluency characterization are therefore decisive in this perspective. What is also new in this thesis is the exploration of the relationship of syntax and prosody in respect of speech disfluencies. Acoustic and prosodic cues have been used for characterizing speech disfluencies, as well as syntactic cues. However, a further and detailed study into the direct mapping of syntactic and prosodic features of the disfluency relations has not been undertaken. The results of the corpus analysis to be presented in later chapters should provide empirical evidence for the production of speech disfluencies, especially from the perspective of syntax and prosody.

If we regard speech disfluencies as integrated units, then the working hypothesis is that they have specific internal relations. By means of empirical studies, evidence is expected to be found. Computational description devices are to be adopted to describe the disfluency relations in a formal manner. In other words, the *regularity* of the syntactic and prosodic features in the production of speech disfluencies is not only empirically observable, but also formally describable.

Chapter 3

Phrase-Based Production of Disfluencies

Linguistic characteristics of disfluencies have been examined in various aspects in the 60s and 70s, including the influential psycholinguistic research work done by Boomer & Dittman [6], Fromkin [29] and Goldman [34]. Attempts at formalizing speech disfluencies in spoken language and improving the results of automatic speech recognition and parsing strategies in computational linguistics have subsequently been made later in the 80s and 90s by Hindle [42], Langer [62], O'Shaughnessy [88], Shriberg *et al.* [112], Nakatani & Hirschberg [82] and Althoff *et al.* [3]. In psycholinguistic fields, one makes use of speech disfluencies in exploring more linguistic evidence of human language processing, whereas computational approaches aim to detect and correct disfluencies in order to run natural language systems successfully. But they more or less all focus on the literally-viewed disfluent components instead of on "an integrated syntax" of the internal structure of disfluent phenomena. For instance, Hindle made use of edit signals and compared the syntactic categories on both sides of edit signals. No specific units were defined. Likewise, Langer applied string-based rules used for detecting syntactic errors without using any explicit units. That is, they regard disfluencies merely as sequences of words, rather than as integrated units.

Disfluencies in spoken language can be pauses, repetitions, speech repairs and speech errors. They all show specific linguistic and communicative peculiarities in the context they occur. From the point of view of mental processing of human language, disfluencies can be alternatively classified into covert repairs, overt repairs and different repairs [5]. Until recently more attention was paid to individual phenomena and the related pragmatic functions of disfluencies than on the structural description of disfluencies. It is not satisfactory to simply annotate disfluencies word by word and not investigate their internal structure. Few approaches have dealt with the internal structure of disfluencies. For example, in [66], repairs are described in terms of *reparandum, editing term* and *alteration* with additional structural information including: *delay* for the rest of the erroneous string after the reparandum and *retracing* for the repeated part in the repaired string before the alteration. Where the problematic string begins and the repaired string ends are in fact not explicitly taken into consideration. Similarly, the other approaches mentioned in Chapter 2 also introduced their own systems of domain definitions of the internal structures of disfluencies. The aim of this chapter, however, is not to give a new definition of internal domains of disfluencies, but to find out what are the internal relations between the domains by investigating real speech data.

Ungrammaticality of spoken utterances does not necessarily imply that syntax does not play a role. In this chapter, empirical support for a phrase-based structural description of disfluencies in spoken language is introduced, based on the results of a corpus analysis focusing on the internal linguistic structure of the production of disfluencies. It is proposed that syntax is still of significant importance in spoken language. The following empirical study should, on the one hand, support the notion that a variety of syntactic features are directly and consistently related to the production of disfluencies in spoken language. On the other hand, the study should also provide evidence for the theoretical consideration of a syntax-based, or more precisely, phrase-based modelling of the internal structure of disfluencies.

This chapter is structured as follows. First, preliminary considerations of the relevance of phrases in the production of disfluencies, especially for disfluencies of complex structure, are briefly presented. Second, definitions used in the oncoming empirical study are given and explained, including definitions for *phrase/disfluency-phrase* and *boundaries of disfluency.* Because the length of a disfluency is to be investigated by counting the number of phrases contained in the disfluency, definitions of phrase and disfluency-phrase are therefore indispensable. Following the introduction on working definitions, corpus data, which are to be examined in the corpus analysis, are presented. Results concerning the syntactic aspects of disfluency production are subsequently presented including

1. *the linguistic length of disfluencies*,
2. *the syntactic category of disfluencies*,
3. *the construction types of disfluencies*,
4. *the location of interruption*,
5. *the repair onset* and
6. *the repair offset.*

Based on the conventional segmentation of repair structure, combination possibilities of *the reparandum*, *the edit signal* and *the repair* are discussed with respect to the internal structure and description of disfluencies. Results obtained by means of the corpus analysis are further specified with the preliminary help of this conventional segmentation of speech repairs. It becomes apparent that the production of disfluencies is frequently oriented toward a phrase-based regularity. Based on the systematic conclusion derived by theoretical consideration of phrases and empirical results of the phrase-based production of disfluencies, a preliminary model of the production will then be introduced.

3.1 Why Phrases?

The relationship between grammaticality and communication in spoken language has created more and more interest in psycholinguistic research fields, as well as in conversational analysis. Taylor & Cameron [120] proposed a target sentence hypothesis:

> The hearer is assumed to have "in store" the units of the language and the possible sentential sequences of such units. So, when he hears an utterance, he attempts to take it as a token instance of one of the stored sentences. During the hearer's decoding of an utterance, then, he is thought to be forming hypotheses as to what sentence-type the "incoming" utterance is a token of. If, having formed one such hypothesis, he can then take the remainder of the utterance as conforming to the predicted sentence-type, he will do so. (Taylor/Cameron (1987), p. 129)

Independent of how different the definitions of sentences in written language and utterances in spoken language are, the unit *phrase* seems to play a combining role. Because units smaller than utterances/sentences and larger than words are phrases. In the case of German, the grammatical coherence within phrases is strongly required. As it is concerned with "units of the language" and "possible sentential sequences", phrases or say phrasal boundaries are supposed to be relevant to the production of utterances in spoken language. In the production of speech repairs, phrasal boundaries also play a role, as mentioned in [66]. The speaker often jumps back to the previous phrasal beginning to initiate his speech repair. That is to say that boundaries of phrasal constituents are highly relevant locations in the production and perception of spoken language.

From the theoretical viewpoint, phrases are getting more and more important in the description of a grammar. It can be observed that recent grammatical theories such as GPSG (Generalized Phrase Structure Grammar) [31] and HPSG (Head-Driven Phrase Structure Grammar) [90]; [91] focus on the role and structure of phrases in describing the grammatical structure of a language, although they start with *sentences* as units. However, it is difficult to recognize and define the boundaries of sentences in spoken language. In contrast, phrasal boundaries can still clearly be identified in spoken language as in written language. Recently, the units "phrases" have been used for parsing as well [1]. It has not been proven in the literature that phrases are the most important syntactic unit in general, concerning language processing and language production. And it is perhaps impossible to do that because different research goals determine different relevance and need in a given field. In the course of this chapter, however, it should be made clear that phrases are important units at least with regard to the speech repairing process. Empirical evidence will be presented in this chapter.

3.2 Working Definitions

Terms such as phrases, disfluency-phrases and disfluencies which are used in the following corpus analysis are defined in this section. Among them, phrases and disfluency-phrases are defined in terms of phrasal rules. It is intended to use them as classification criteria to investigate the distribution of the linguistic categories of disfluencies. For a similar reason, boundaries of disfluencies are given in order to determine the length of disfluencies. This section does not attempt to construct a sophiscated grammar structure, but rather to build simple heuristic criteria in order to enable the identification and classification of disfluencies.

3.2.1 Phrase and Disfluency-Phrase

The term *Phrases* used in this chapter is temporarily defined according to conventional grammatical rules. Because in the empirical analysis simple and clear categorizations are required to provide a better overview of the phenomena examined. Theoretical considerations and linguistic modelling of syntactic characteristics will not be introduced until later. They will be thoroughly discussed in chapters specifically dealing with syntactic characterizations. Relevant grammatical rules used for the classification of grammatical categories of disfluencies are DCG-like (Definite Clause Grammar) grammar rules. An utterance (U) is a composition of NP, PP[1], VP, ADV, OTHER and disfluency-phrases (DFP). NP includes pronouns and nouns with or without adjectives, whereas there is no NP with attached prepositional modifiers allowed, such as *the little girl with a book*.

Furthermore, because it is spoken language that these phrasal rules are applied to, NP rules in this system include extra types, which are constructions used very frequently in spoken German. **So** in the construction *so DET (ADJ) N* refers to specific subjects. This construction can be used as an anaphoric reference[2] [25]; [23] or as reference for objects which have not been specified earlier[3]. PP can be divided into preposition and NP. In contrast to conventional predicate definition, the VP category used here is a verb itself with no NP attached. In the case of German, suffixes and participles belonging to verbal constructions are included in the VP as well. Adverbs such as *genau* (exactly) and *jetzt* (now) are defined as ADV. OTHER are conjunctions and unidentified word fragments which can not be identified as belonging to any of the phrases above. Relative clauses which function as noun phrases in German, are considered in terms of their phrasal sub-parts, instead of as independent NP. The grammar rules are listed below as follows:

$U \rightarrow [NP|VP|PP|ADV|OTHER|DFP]^{+}$
$NP \rightarrow N$

[1] In order to keep the grammar rules as flat as possible, I exclude NP rules with PP attachments. NP + PP are treated as two separate phrases.

[2] It is derived from *solch ein* which means *such a*.

[3] This form is derived from *irgendso ein*, meaning *something*.

$NP \rightarrow DET + ADJ^* + N$
$NP \rightarrow DET + ADJ^*$
$NP \rightarrow so + DET + ADJ^* + N$
$NP \rightarrow so + DET + ADJ^*$
$PP \rightarrow P + NP$
$VP \rightarrow V$
$ADV \rightarrow ADVERB$
$OTHER \rightarrow CONJUNCTION$
$OTHER \rightarrow UNIDENTIFIED\ WORD\ FRAGMENT$

where U is an utterance and NP, VP, PP, ADV, OTHER and DFP can appear in any number and in any order. X^+ means that X must appear at least once, whereas X^* means that X can be empty and can occur more than once.

As we aim to deal with disfluent phenomena in spoken language, we also need phrasal rules for the interrupted or incomplete (excluding ellipses) phrases. DFP is represented in the notation X_d, where X is one of NP, PP, VP and ADV. When there is at least one element within a phrase which is not completely produced or the element is erroneous, this phrase fragment or erroneous phrase is a DFP.

$NP_d \rightarrow N_d$
$NP_d \rightarrow DET_d$
$NP_d \rightarrow DET_d + ADJ_d^* + N_d$
$NP_d \rightarrow DET + ADJ_d^* + N_d$
$NP_d \rightarrow ADJ_d^* + N_d$
$NP_d \rightarrow so + DET_d$
$NP_d \rightarrow so + DET_d + ADJ_d^* + N_d$
$NP_d \rightarrow so + DET + ADJ_d^* + N_d$
$PP_d \rightarrow P + NP_d$
$PP_d \rightarrow P_d + NP_d$
$VP_d \rightarrow V_d$

$ADV_d \rightarrow ADVERB_d^*$

where X_d means that X is a word or phrasal fragment or X contains problem elements.

When the length of disfluencies is mentioned, a disfluency-phrase is counted as having the length of one phrase. For instance, the disfluency *the re the red book* has the length of two phrases. An explicit and specific definition of "sentence" is not necessary in this study, because only local disfluencies (speech errors or immediately produced speech repairs) are considered. Disfluencies which are related to longer units such as sentences, moves or turns are not of main interest in this thesis.

3.2.2 Disfluency Boundaries

This section is concerned with the criteria of identifying disfluency boundaries in the syntactic analysis. Before disfluency boundaries are defined, some definitions of disfluencies in the literature will be first summarized.

Levelt [66] dealt exclusively with speech repairs and described repair structures by the reparandum, the alteration, the retracing and the delay (cf. Section 2.2.1). Hindle divided disfluencies into three groups: unusual constructions, true ungrammaticalities and self-corrected strings [42]. Nakatani & Hirschberg [82] distinguished three intervals within speech repairs: the reparandum interval, the disfluency interval and the repair interval. This segmentation is principally based on Levelt's segmentation. Shriberg [110] and Lickley [73] both regarded pauses, repetitions and repairs as disfluencies. They both adapted Levelt's segmentation for describing the internal structure of speech repairs.

In this thesis, disfluencies are word/word fragment sequences in spoken language whose structural occurrence does not satisfy the syntactic rules given for a specific language or they are word/word fragments within the sequences which can not be found in a given lexicon of the language. Therefore, they include pauses, repetitions, speech errors and speech repairs.

Before the length of disfluency is defined, two terms are to be introduced, namely: problem and corrected items. A problem item is a word sequence which can neither be found nor accepted according to the language-specific lexicon and grammar. Pauses and interjections are for instance problem items, as well as interrupted words or phrasal fragments[4]. Corrected items are identifiable [5] sequences which are produced after the problem items to get them repaired. For instance, in the sequence *die ro äh die rote Schraube*[6], the problem item is *ro* and the corrected item is *rote*.

The length of disfluencies is thus defined as follows: the beginning of a disfluency is the nearest phrasal boundary before the identified problem item and the end of a disfluency is the end of the next phrasal boundary after the identified corrected item, when available. If there are no corrected items, the end of the disfluency is then the next phrasal boundary directly after the problem item. Figure 3.1 illustrates this definition. The interval [**a**, **b**] is a disfluency, where **a** is the nearest phrasal boundary before the problem item and **b** is the nearest phrasal boundary after the corrected item.

The examples below should make this definition clearer. The underlined sequences are disfluencies, if the above definition is applied. In Example 3.1, *eine gel eine gelbe Schraube* is a disfluency with the problem item *gel* and the corrected item *gelbe*, whereas in *mit äh*, the problem item is *äh* and the phrasal boundaries are those before *mit* and after *äh*. Simple disfluencies

[4]Phrasal sequences such as *von der* (from the) or *auf die* (on the) are problematic, because it can be ellipse as well. In such cases, the contextual information is then required to check whether they are problem items or they are simply ellipses.

[5]Some speech repairs which aim to correct the previous errors can be incomplete.

[6]the re uh the red bolt

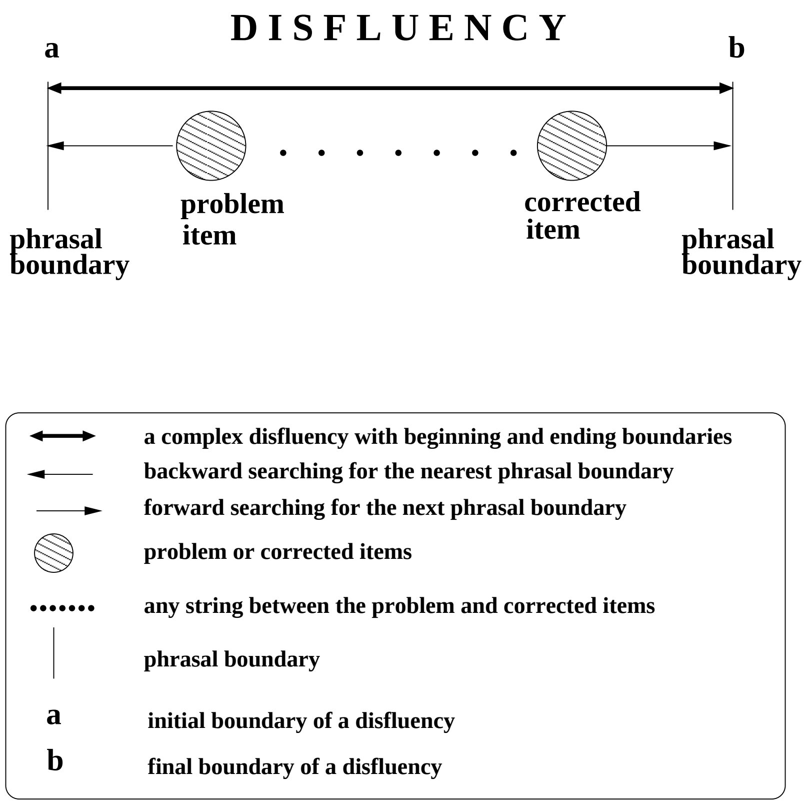

Figure 3.1: Definition of Disfluency Length

can also be identified using this definition. As shown in Example 3.2, *öh* is realized between two adverbs, i.e. *öh* is located between two phrasal boundaries. *Eine rund Schl* is a speech error without being repaired, so the length of this disfluency is from the phrasal boundary preceding *eine* to the position where it is interrupted. Repairing can also take place within a lexical entry as shown in Example 3.3. So the problem item is *Dreierlis* and the corrected item is *leiste* in the disfluency *die Dreierlis äh leiste*. Althoff *et al.* [3] applied the word lattice method in dealing with compound words containing this type of disfluency.

Example 3.1
dann eine gel eine gelbe Schraube mit äh eine runde mit Schlitz
(then one ye one yellow bolt with uh one round one with a slot)

Example 3.2
dann öh nochmal eine achtkantige rote und nochmal eine runde Schl
(then uh again one eight edge red one and again one red sl)

Example 3.3
Ich habe die Dreierlis äh leiste noch frei
(I have the three-holes l uh bar still free)

3.3 BAUFIX Corpus

Corpus analysis is regarded as a useful method in recent linguistic research fields. Data and results obtained by means of corpus analyses can provide empirical support for proving or disproving linguistic theories. The corpus used in this thesis is the BAUFIX corpus. Applying the definitions given earlier, disfluencies in BAUFIX are identified and are subsequently analysed in terms of their syntactic features.

3.3.1 Corpus Data

The BAUFIX corpus [97] has been made available within the framework of the project SFB 360 (Sonderforschungsbereich 360) at the University of Bielefeld. It contains 22 instruction-construction spoken dialogues. Constructors had to co-operate in building a toy plane following the instructions given by their partners: instructors. Instructors had either a model or a sketch of the toy plane to help in formulating their instructions verbally. Subjects were students from the University of Bielefeld. In 9 of the 22 dialogues, subjects could not see each other at all and had to rely solely on their verbal formulations and communication. In 6 dialogues, there was only limited visual contact between the interlocutors. In 7 dialogues, subjects had free visual access to each other's movements and actions. These task-oriented spoken dialogues were digitally recorded at the sampling rate of 44.1 kHz and were orthographically transcribed. The conversations are therefore situated spontaneous speech which nevertheless contains a variety of spoken language phenomena, i.e. disfluencies. Here are some basic data on the corpus:

number of dialogues:	22
male subjects:	15
female subjects:	29
number of lexical items uttered:	35036

3.3.2 Disfluencies

Applying the rules given above with the help of an orthographic transcription and the DAT (*digital audio tape*)-recordings, disfluencies are identified. Filled and silent pauses occurring in syntactically correct context are also counted as disfluencies in spoken language. However, they will not be taken into account here, for reasons that they only provide very limited information about the internal structure of complex disfluencies and therefore are not the main interest in this thesis, namely the internal relationships within disfluencies. Likewise, interjections and pauses such as *the uhm red bar* or *the :p: red bar*[7] are not taken into account in the corpus analysis, nor are simple disfluencies which are produced immediately between two phrasal boundaries. In total 500 disfluencies were identified in the 22 dialogues. They can be found in Appendix C.

3.4 A Syntactic Analysis of Disfluencies

In this section, a syntactic analysis of disfluencies found in the BAUFIX corpus is carried out. Syntactic features are examined focusing on linguistic length, category and construction types as well as presumably relevant positions within and around disfluencies are examined. As speech repairs can provide more specific and complex information about the internal structure of disfluencies than other types of speech disfluencies, more investigation into the structural characteristics of speech repairs is carried out.

3.4.1 Linguistic Length of Disfluencies

Because one of the aims of this syntactic analysis is to investigate the relationship between the syntactic unit *phrase* and disfluencies, 500 identified disfluencies are classified into three groups with respect to their linguistic length, namely lengths of one word, of one phrase and of more than one phrase.

I. disfluencies covering only one lexical item, including completed, interrupted or erroneous items

Examples:

- ja, *wievi*, na ja, alles klar[8].

[7] :p: stands for silent pauses in this thesis.

[8] how *mu* well it's ok

- auf der linken Seite dann *die*[9].

II. disfluencies covering only one phrase, including completed, interrupted or erroneous phrases

Examples:

- und oben *die ro*, ja[10]
- *diese zwei orangenen äh violetten ringe*[11]

III. disfluencies covering more than one phrase

Examples:

- *Das, äh der* ist ja frei[12].
- *auf diesen auf diese Dreierleiste* setzt du jetzt eine Fünferleiste[13]
- wenn man es so *von von den Sch* auf die Schrauben guckt[14].

Table 3.1: Disfluencies in Syntactic Context

category	frequency	percentage
I.	54	10.8%
II.	84	16.8%
III.	362	72.4%
total	500	100%

Table 3.1 shows the structural characteristics of disfluencies from the point of view of their syntactic length. Disfluencies whose length are of only one lexical item make up about 11%. Similarly, about 17% out of 500 disfluencies were realized within one phrase. The majority of disfluencies,

[9] on the left hand side then *the*

[10] and above *the re*

[11] *these two orange äh violet rings*

[12] *the*[neuter, singular, nominative, definite] *äh the*[masculine, singular, nominative, definite] is free

[13] *on this on this three-holes bar* you now put one five-holes bar

[14] when one looks it so *from from the b* onto the bolts.

72%, consist of more than one phrasal length. This result clearly suggests that a complex structural description of disfluencies requires a linguistic unit larger than lexical items, since the majority of disfluencies have the length of at least one or more than one phrase, as illustrated in Figure 3.2. This empirical result preliminarily supports the hypothesis that the unit *phrase* is a possible suitable unit of describing the structure of disfluencies.

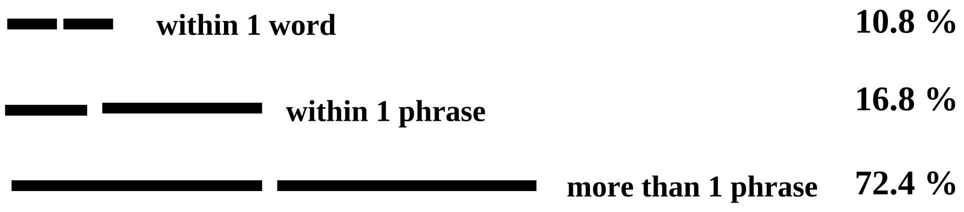

Figure 3.2: Linguistic Length of Disfluencies

3.4.2 Syntactic Category of Disfluencies

In which linguistic categories disfluencies often occur is closely related to the specific type of corresponding spoken discourse. This chapter proposes that the syntactic unit, *phrase*, is a suitable unit for the structural description of disfluencies. Thus, this section investigates how the production of disfluencies is related to their syntactic categories in instruction-construction task-oriented dialogues. 500 disfluencies are classified into five grammatical categories: prepositional, noun, verb, adverb phrases and other-category, depending on which category a disfluency was produced in. They are annotated by PP, NP, VP, ADV and OTHER for convenience as those used in Section 3.2.1. But PP, NP, VP, ADV and OTHER used in this section exclusively represent the categories of disfluencies where they are located and are not related to the phrasal structures of disfluencies. Thus, they can not account for disfluencies containing more than two phrasal types. An additional category is therefore added to the classification, namely 3P. 3P are disfluencies covering three or more than three phrases. In fact, there are very few disfluencies containing more than three phrases, because only immediate speech repairs are considered disfluencies in this syntactic analysis. Speech repairs across utterance levels are not taken into consideration. The result are presented in Figure 3.3:

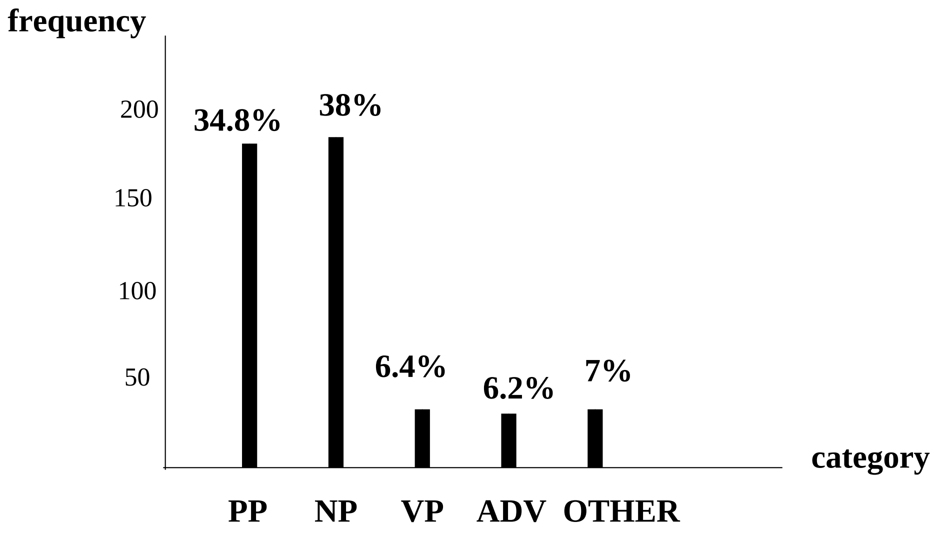

Figure 3.3: Disfluencies in Syntactic Category

Prepositional and noun phrases are obviously the most frequent phrases, in which errors and repairs are produced. They make up about 75% of the disfluencies in the 22 dialogues. The corpus contains instruction-construction dialogues in which the instructors have to describe objects: their *type*, *size* and *colour* and the steps of constructions, *direction* and *location*. Prepositional and noun phrases are exactly the phrases whose contents give this type of information and they play a central role in language usage under this specific experimental condition. More specifically, the subjects had to describe the objects and the circumstances of the objects explicitly to fulfill their task, namly the construction of the toy plane.

These result also provide additional insight into the relationship between the length and the categories of disfluencies. Although complex disfluencies such as speech repairs are supposed to be longer than other types of disfluencies, 3P disfluencies such as *und eine und eine Mutter* [15] are not the majority. It accounts for merely 7.6% of the disfluencies. Therefore, one can conclude that the number of phrases between the problem and corrected items is seldom larger than two in most cases.

As prepositional and noun phrases account for about 75% of all disfluencies found in the corpus, further investigations of the structural features of disfluencies in the coming sections will mainly concentrate on the disfluencies produced within these two phrases.

3.4.3 Construction Types of Disfluencies

In order to explore more structural characterizations of disfluencies realized in prepositional and noun phrases in German, the most frequent types of syntactic constructions of disfluencies are to be investigated in this section. Disfluencies in prepositional phrases are divided into five types of phrasal constructions, as illustrated in Figure 3.4. According to the surface structure of disfluencies and the environmental context of disfluencies (including their later correction), they can be classified into groups from PP_1 to PP_5. Disfluencies which can not be identified as any one of the categories mentioned above are classified into the PP_rest group such as phrasal fragments like *mit äh mit* [16]. Because this phrase is not yet finished and it is not clear which syntactic construction the speaker wants to use, this kind of phrase fragment is classified into the PP_rest group.

The distribution of the occurrence of 174 disfluencies in the prepositional phrases is shown in Figure 3.5 and Table 3.5. The results show, based on the constructions which disfluencies in the prepositional phrases have, that disfluencies in prepositional phrases are often produced in the form of $PP => P + DET + N$ and $PP => P + DET + ADJ + N$. These two constructions

[15] and one[feminine, singular, indefinite] and one[feminine, singular, indefinite] nut

[16] with äh with

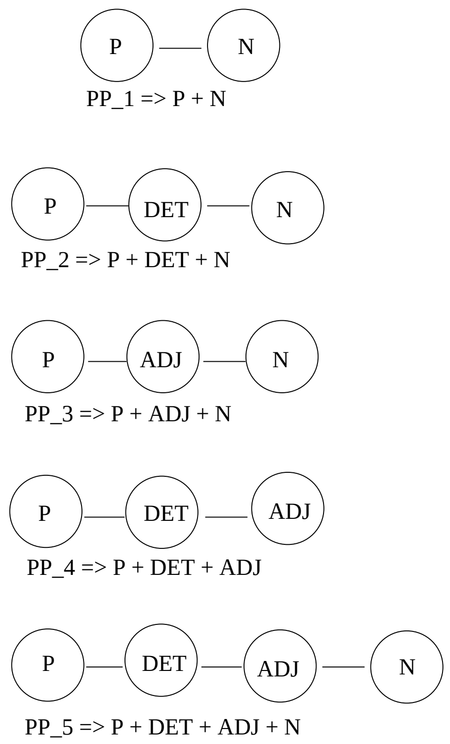

Figure 3.4: Prepositional Phrase Constructions

account for more than half of disfluencies in the prepositional phrases found in the corpus. It is also possible that disfluencies found in the rest category may belong to these two constructions mentioned above.

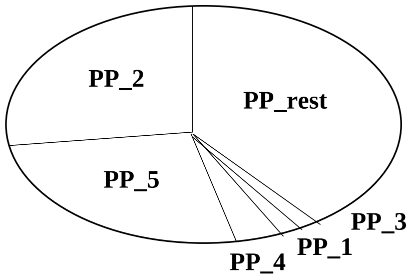

Figure 3.5: Distribution of PP Constructions

Table 3.2: Phrasal Construction in PP

construction type	**frequency**
PP_1	5
PP_2	46
PP_3	4
PP_4	13
PP_5	47
PP_rest	59
total	174

The same investigation is carried out for disfluencies produced in the noun phrases. The distribution of the six phrasal construction types of disfluencies in the noun phrases is illustrated in Figure 3.6. N stands for both pronouns and nouns in this case. Unidentifiable noun phrase fragments are, as in the prepositional cases, classified into the rest group NP_rest.

The distribution of disfluencies in the noun phrases is presented in Figure 3.7 and Table 3.7. Exactly as in the prepositional cases, disfluencies are

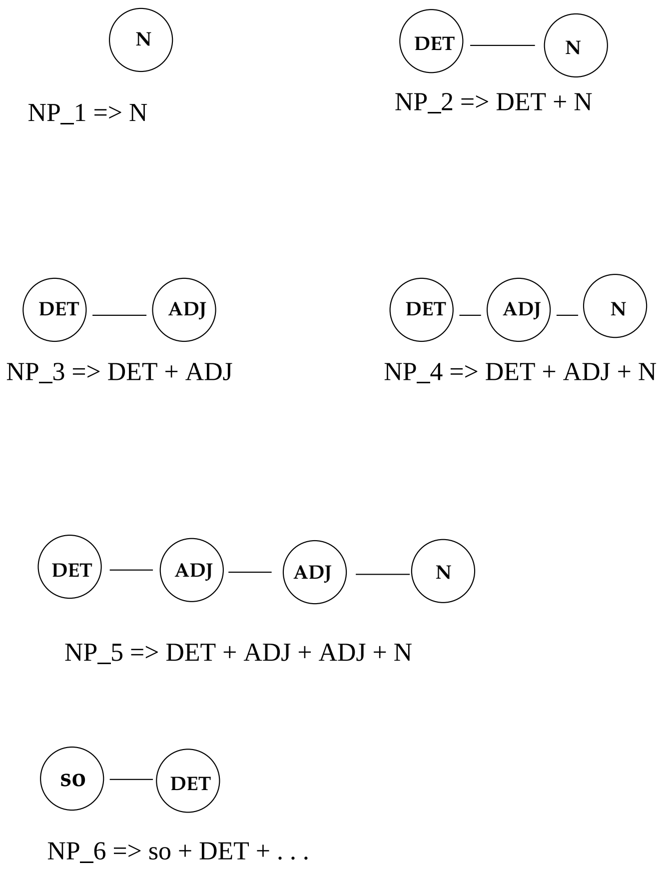

Figure 3.6: Noun Phrase Constructions

much more often realized in certain constructions, namely $NP => DET+N$ and $NP => DET + ADJ + N$. As shown in the results, there clearly exists a relationship between syntactic structure and the occurrence of specific type of disfluencies. The reason being that disfluencies are more frequently realized in certain syntactic environments and contexts than in the others.

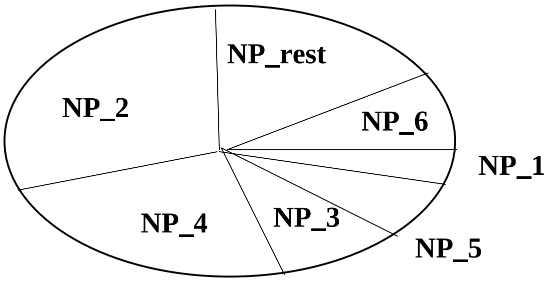

Figure 3.7: Distribution of NP Constructions

Table 3.3: Phrasal Construction in NP

construction type	**frequency**
NP_1	5
NP_2	51
NP_3	19
NP_4	49
NP_5	11
NP_6	12
NP_rest	43
total	190

3.4.4 Location of Interruption

In most of the literature on speech repair research, the location of interruption is an important cue in detecting disfluency as well as for correcting speech errors and therefore detecting speech repairs. In this section, the position of interruption is to be investigated among disfluencies identified within prepositional and noun phrases. Location of interruption is defined as the position where the speaker interrupts his contribution which can be identified by means of pauses, word fragments or some lexical indicators such as *well* or *you know*.

As shown in Figure 3.8 and 3.9, when word fragments occur in prepositional phrases, they are always located in adjectives and nouns, but never in determiners or prepositions themselves. Interruptions after word boundaries are located in the position after determiners in 44% of the prepositional phrase disfluencies. In about 25% of prepositional phrase disfluencies, the interruption is positioned after prepositions. This result is indicative of the peculiar feature of German articles which differentiate three grammatical genders. The special role German articles play becomes clearer when we observe the location of interruption of the noun phrase disfluencies in Figure 3.10 and 3.11. In 63% of the noun phrase disfluencies, the interruption comes after the determiners. It is also worth mentioning that word fragments occur more often in adjectives than in nouns.

3.4.5 Two Results of Preliminary Studies

Two important results obtained in the studies until now will be briefly summarized in this section, because they will be applied to the disfluency modells in later chapters. First, the majority of disfluencies in prepositional and noun phrases contain determiners. Determiners were found in 72% of the prepositional phrase disfluencies, whereas 75% of the noun phrase disfluencies had determiners. Second, if there are determiners in the prepositional and noun phrase disfluencies, it is very frequently interrupted after the determiners. 81% of the prepositional phrase disfluencies containing determiners were interrupted after the determiners, whereas interruptions were located after the determiners in 85% of the noun phrase disfluencies containing determiners.

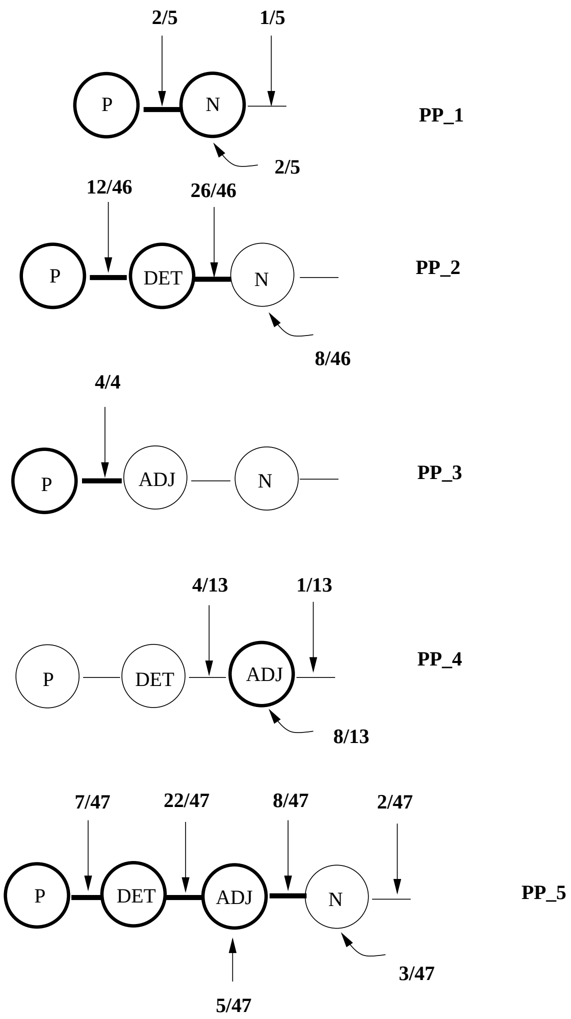

Figure 3.8: Location of Interruption in PP_1-PP_5

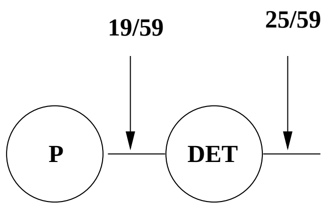

Figure 3.9: Location of Interruption in PP_rest

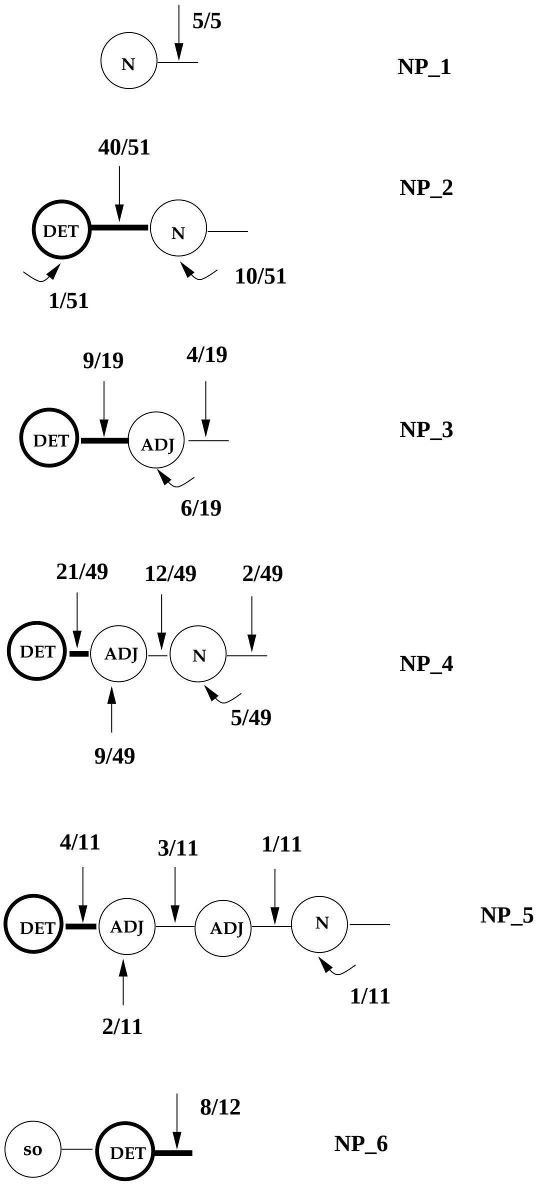

Figure 3.10: Location of Interruption in NP_1-NP_6

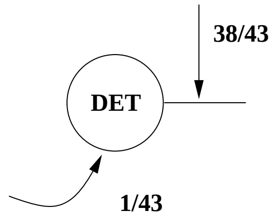

Figure 3.11: Location of Interruption in NP_rest

It is a significant result: when interruptions occur in prepositional and noun phrases, they are usually realized after determiners.

3.4.6 Repair Onset

Not all disfluencies are produced in the form of speech repairs. Some of them are simply interruptions or speech errors without being repaired. This section deals with speech repairs and, in particular, how corrected items are initiated will be the focus of the investigation. The phrase in which the corrected items occur is defined as the *corrected phrase*, whereas the phrase containing problem items is called the *problem phrase*. The position of the onset of the corrected phrase is examined for disfluencies in prepositional and noun phrases. The results are illustrated in Figure 3.12 and 3.13. It is clear that the corrected phrases in most cases were initiated accordingly at the beginning of the previous problem phrase, both in prepositional and noun phrases.

120 out of 174 prepositional phrase disfluencies and 149 out of 190 noun phrase disfluencies were repaired by initiating the repairing process at the beginning of the previous phrasal boundary. This result clearly shows that phrasal boundaries are directly related to the production of speech repairs. Based on this result, it is possible that concerning the production of speech repairs, the number of retraced syllables of speech repairs regarding the phenomenon of *retracing*, as suggested by Levelt [66], may be less important

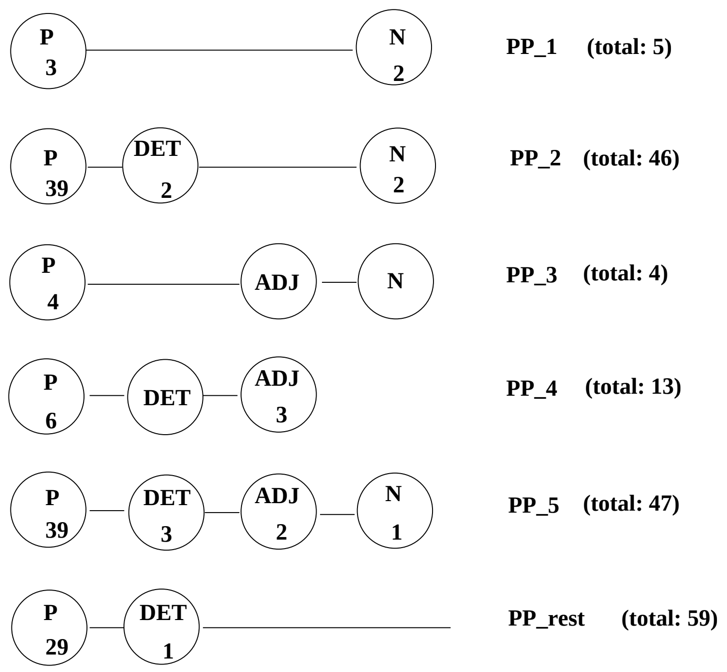

Figure 3.12: Position of Repair Onset in PP

than the position of the phrasal boundary immediately preceding the speech error.

3.4.7 Repair Offset

This section aims to examine the repair offsets located within prepositional and noun phrases. The repair offset is defined to be the position where the corrected phrase ends. Whether the last word in the corrected phrase is completed or not does not play a role here, because it is possible that the speaker does not complete his/her repair in the first attempt. Results in Figure 3.14 and 3.15 show that the corrected phrases are usually completed phrases, i.e. repairs are realized as complete phrases without ellipses. It is seldom the case that only the correction of the problem item is realized, even if the problem item is located in an early position within the phrase. Instead,

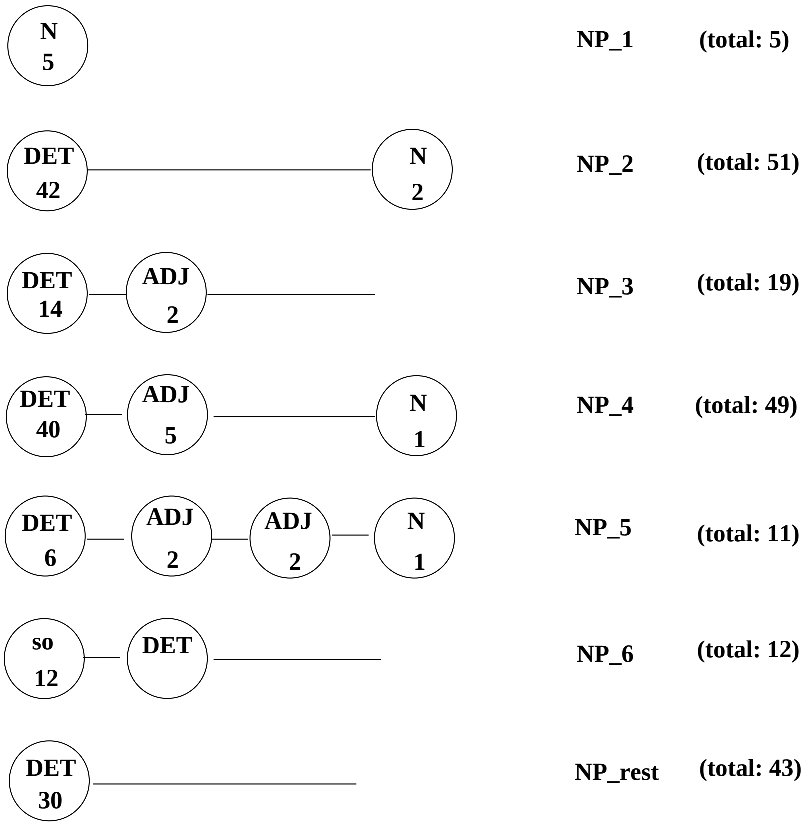

Figure 3.13: Position of Repair Onset in NP

the majority of speech repairs end in the phrasal-final position. 97 out of 174 prepositional phrase disfluencies and 102 out of 190 noun phrase disfluencies are repaired in terms of this construction[17]. The reason for the smaller percentage in both cases, compared with that in the case of repair onset, is that in the rest category, PP_rest and NP_rest, many of the disfluencies are simply interrupted phrase elements and not repaired.

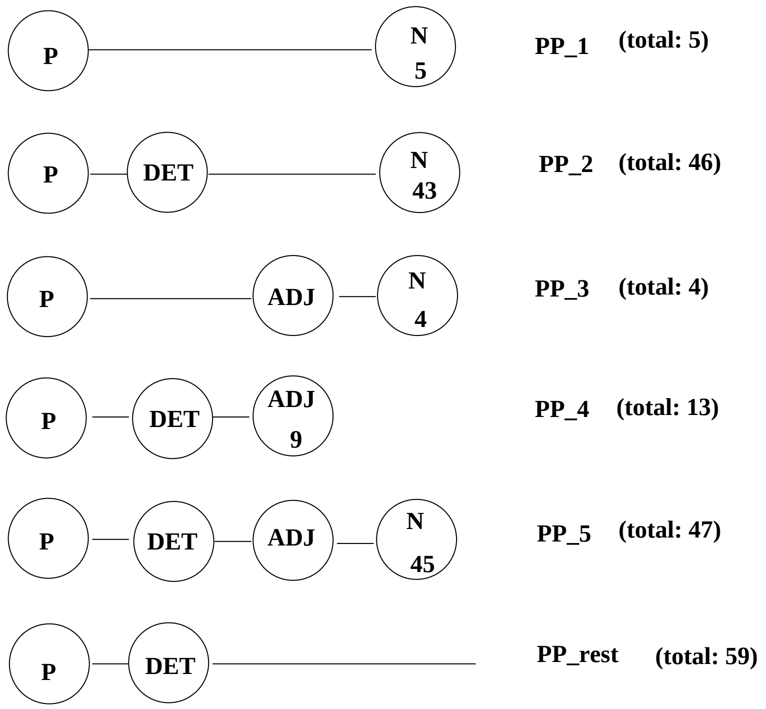

Figure 3.14: Position of Repair Offset in PP

In [66], the relevance of phrasal boundaries has been mentioned, however, only the phrasal boundaries related to the retracing phenomenon. The position of where a corrected phrase ends has not explicitly been taken into consideration, although, as shown here, it holds a highly relevant position

[17]The reason that NP_6 is not included in this calculation is that the construction types of the 12 NP_6 cases need a new sub-classification and the sub-groups are much smaller than the other NP categories. Therefore, they are not included in the figure and in the calculation.

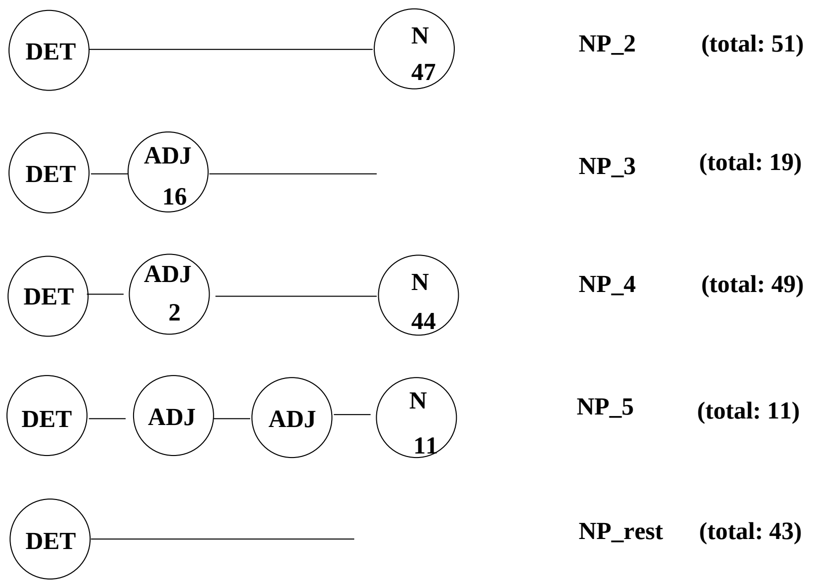

Figure 3.15: Position of Repair Offset in NP

among the speech repairs investigated.

3.5 Reparandum, Edit Signal and Repair

This section deals with the relationship between the conventional segmentation of speech repairs, namely, reparandum, edit signal and repair, and the classification of disfluencies. In terms of these three parameters, one can combine them to form possible types of speech disfluencies, as illustrated in Table 3.4.

In order to make clear how disfluencies can be structured, examples are given for each combination in this section. Type I is the "full form" of disfluency occurrences where the reparandum, the edit signal and the repair can be easily identified. The relationship between the reparandum and the repair functions as a *replacement*. The forms of replacement can be substitution, deletion or addition [13].

Table 3.4: Types of Disfluencies in Spoken language

	Reparandum	**Edit Signal**	**Repair**	**Disfluency Types**
I:	+	+	+	speech repairs
II:	+	+	-	speech errors
III:	+	-	+	speech repairs
IV:	+	-	-	speech errors
V^*:	-	+	+	
VI:	-	+	-	pauses/ lexical markers
VII^*:	-	-	+	
VIII:	-	-	-	

Example 3.4
Und unten runten ist halt die gelbe Mutter äh die orange Mutter. (Sagerer et al. 1994)
(And beneath that is the yellow nut eh the orange nut.)

Example 3.5
Dann hast Du die braunen äh die blauen langen Schraube. (Sagerer et al. 1994)
(Then you have the brown eh the blue long bolt.)

Type II is not difficult to identify, since the utterances containing type II are never well-formed, are often interrupted and contain edit signals. This kind of disfluency is realized as a speech error.

Example 3.6
und nimm ähm Du hast diese zwei äh langen Latten noch nicht angebracht ne? (Sagerer et al. 1994)
(and take uhm you haven't fixed these two eh long slats yet, right?)

In Type III, the reparandum and the repair are realized without edit signals (Example 3.7). This combination shows that the interruption between the reparandum and the repair sometimes doesn't take place at all. The temporal interval, whether it is filled or not, between the reparandum and the repair may not even exist. The same result was also obtained by Blackmer & Mitton [5].

Example 3.7
Du nimmst jetzt die orange, längs öh Schritz Schlitzschraube. (Sagerer et al. 1994)
(Now take the orange, long eh srot slot bolt.)

Example 3.8
Ja, zusammen mit der mit der mit den drei Löchern. (Sagerer et al. 1994)
(Yes, together with that with that with the three holes.)

The combination in that speech errors are produced without edit signals and corrections is type IV. Utterances are simply interrupted, after speakers detect their errors.

Example 3.9
Das ist bei mir auf der rech, ich kann es auch umdrehen. (Sagerer et al. 1994)
(It is for me at the righ, I can also turn it.)

Type V and VII should not exist. If the reparandum can not be identified, no repair can be identified either. Type VI appears as interjections which are inserted within phrases or across phrasal boundaries.

Example 3.10
Und ähm nimmst Du ähm den, ja, nimmst Du die gelbe Schraube mit der Kerbe. (Sagerer et al. 1994)
(And uhm you take uhmthe, well, you take the yellow bolt with the notch.)

Type VIII is seldom expected in an unplanned talk, in which no disfluencies occur. Usually, only trained speakers can achieve delivering verbal information fluently without noticeable corrections during the flow of speech. It is especially difficult not to use interjections during spoken language as the following sentence shows:

Example 3.11
Ich habe nämlich vergessen, daß Du jetzt noch eine Dreierstange hast und die mußt Du da noch zwischenbauen. (Sagerer et al. 1994)
(I forgot, that you still have another three-holes bar that you have to screw on first.)

3.6 Problem Phrase, Editing Phase and Corrected Phrase?

In terms of the combinations listed in the previous section, reparandum, edit signal and repair, each can represent different types of disfluencies. As shown in the corpus analysis above, phrasal boundaries around disfluencies mark most of the important sites of disfluencies. Investigations of the syntactic length and the category of the disfluencies identified, also demonstrate that phrases play an organizing role in the production of disfluencies. Based on the empirical data previously obtained, it can be concluded that the reparandum usually coincides with the problem phrase, whereas the repair is closely linked with the corrected phrase.

For most cases found in the corpus, the segmentation of reparandum, edit signal and repair for complex disfluencies can be specified by problem phrase, editing phase and corrected phrase. Table 3.5 shows the number of disfluency cases found in the BAUFIX corpus which can be described in terms of the specification (problem phrase, editing phase, corrected phrase), by varying the parameters. 88% of PP disfluencies and 86% of NP disfluencies identified in the corpus can be described in this form.

Table 3.5: Specified Types of Disfluencies in Spoken language

Disfluencies	**Form**	**PP**	**NP**
repetitions/	(problem phrase, _, corrected phrase)	6	25
repetitions/ restarts	(problem phrase, editing phase, corrected phrase)	1	1
pauses	(_, editing phase, _)	-	-
speech errors	(problem phrase, _, _)	28	14
	(problem phrase, editing phase, _)	9	6
speech repairs	(problem phrase, _, corrected phrase)	96	97
	(problem phrase, editing phase, corrected phrase)	13	21
sum		153	164
total		174	190

3.7 Phrase-Based Production

In this section, a preliminary model of the production of disfluencies in spoken language is proposed. The relevant results concerning the relationship between phrasal boundaries and the internal structure of disfluencies are systematized and specified to obtain an overall picture of the model. Three important positions: *error onset*, *location of interruption* and *repair offset*, are primarily considered for a structural description of disfluencies. As shown in Figure 3.16, complex disfluencies should be considered integrated units, not individual and separated phenomena such as speech errors, pauses, or speech repairs. Since speech repairs across phrasal boundaries represent the majority in the corpus data, most of this kind of complex disfluencies can be modelled as in Figure 3.16. Error onset and repair offset which were mentioned in Figure 2.6 of Section 2.2.1 can now be clearly defined.

The interval [**a**, **b**] in Figure 3.16 is defined as a disfluency, whereas [**c**, **d**] is the editing phase as defined in [66] which are elements occurring between errors and repairs. [**a**, **c**] represents the problem phrase where the error is located and [**d**, **b**] is the corrected phrase where the repair is realized. Not

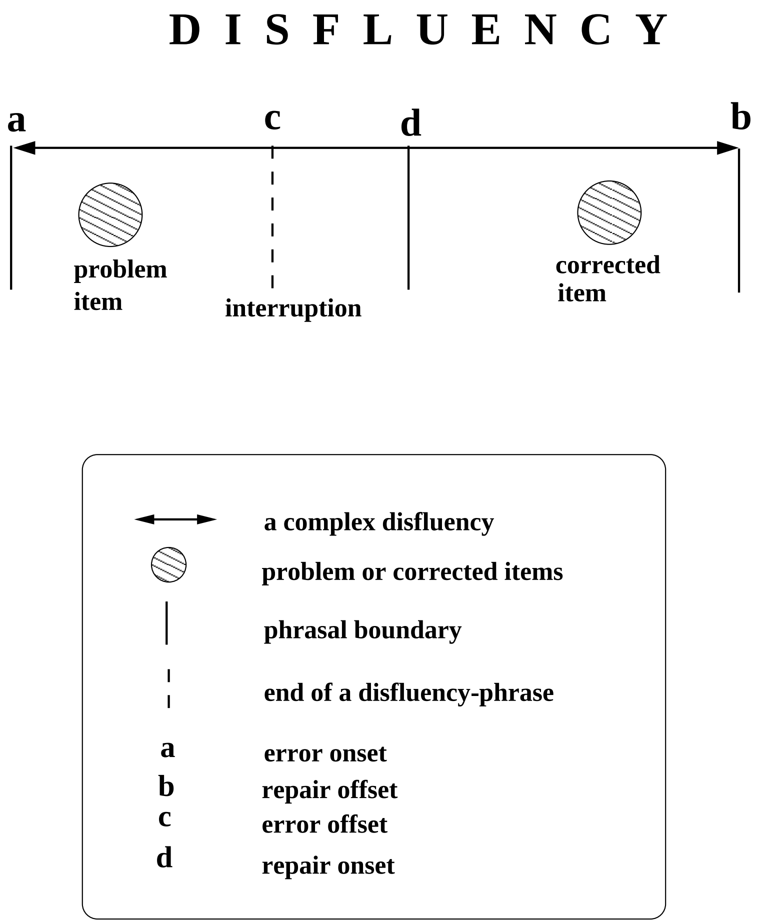

Figure 3.16: Phrase-Based Description of Disfluencies

all the intervals between **a** and **b** are supposed to be present. For instance, a pause is to be described in terms of the interval [**c**, **d**] without [**a**, **c**] and [**d**, **b**]. An interval [**a**, **c**] is then simply a speech error without being repaired. It certainly does not make sense to have [**d**, **b**] available and [**a**, **c**] absent, because a repair can not be existent when no error can be identified. The structural characterizations of disfluencies illustrated in Figure 3.16 will be presented in Section 3.7.1, 3.7.2 and 3.7.3.

3.7.1 Error Onset and Error Offset

As shown in Figure 3.16, error onset and error offset are identified as follows.

error onset = start of the problem phrase

The position **a** is the phrasal boundary immediately preceding the problem item. In other words, it is the beginning of the problem phrase. So, the error onset is defined as the start of the problem phrase.

error offset = end of the problem phrase

In Figure 2.6, we can clearly identify the location of the error offset, namely the location of interruption. This corresponds to **c** in Figure 3.16, where **c** is not only the location of interruption but also the end of the problem phrase.

3.7.2 Repair Onset and Repair Offset

Similarly, repair onset and repair offset are also defined as illustrated in Figure 3.16.

repair onset = start of the corrected phrase

The phrase in which the corrected item occurs is in most cases a new phrase, namely [**d**, **b**] in Figure 3.16, instead of a continuation of the phrase containing the problem item. After the problem item or interruption, one tends to initiate his repair with a new phrase. Therefore, in the results in Section 3.4.1, only few disfluencies are shorter or longer than two phrases. Similar results are also shown in Section 3.4.2 that disfluencies longer than two phrases are in the minority.

The phenomenon of repair retracing defined in [66] can be empirically re-examined by means of the empirical data obtained in the corpus analysis. The result of the corpus analysis strongly suggests that it is more useful to look at the syntactic position where disfluencies are initiated than to calculate a numerical repair retracing value as Levelt did. Because the distance of retracing can be extremely variable it is therefore very difficult to obtain systematic control over the retracing behaviour.

Repair Offset = End of the Corrected Phrase

Corrected phrases are usually complete phrases. It is seldom the case that problem items are corrected without repeating the initial parts of the problem phrase, if the problem items are not in phrasal-initial positions. Speech repairs often appear as a *corrected phrase*, the interval [**d**, **b**] in Figure 3.16, instead of *corrected words*. Using the definition of corrected phrases, we can define the repair offset as the end of the corrected phrase.

Where the phrase containing corrected items ends is seldom where the phrase containing the problem item has ended. Instead, phrases containing corrected items are usually complete phrases. That means after the errors are corrected, the speaker still utters the phrase to the very end of the phrase. Seldom are ellipses used in this case.

3.7.3 Error Onset vs. Repair Onset

In most of the complex disfluencies in prepositional and noun phrases, the onset position of the problem and corrected phrases, in Figure 3.16 **a** and **b** respectively, is identical with respect to their phrasal structure and part of speech information. This fact points out the tendency that retracing phenomena often takes place in the case of speech repairing and the position where it takes place is usually consistent.

3.8 Summary

In Chapter 3, a corpus analysis was carried out to investigate syntactic characteristics of speech disfluencies. Starting with preliminary definitions of disfluency-phrase and disfluency boundaries, interesting results were obtained. Most of the complex speech disfluencies were produced in terms of very consistent syntactic structure. Complex disfluencies produced within prepositional and noun phrases were far more often found than within other categories. The length of a complex speech disfluency, often called a speech repair, is in most of the cases not longer than a disfluency-phrase plus a phrase. The first "half" of a complex disfluency, often a disfluency-phrase, corresponds to the definition of the repairable part, called the reparandum in the literature, whereas the second "half" of a complex disfluency, usually a complete phrase, is the repair. That is to say, the rather general segmentation of speech repairs, "the reparandum and the repair", is realized in real speech material in a specified form: "a disfluency-phrase and a corrected phrase". Furthermore, the construction types of disfluencies produced in prepositional and noun phrases were not realized in an arbitrary way. These are certain preferred constructions. For instance, in prepositional phrases P + DET + N and P + DET + ADJ + N were used more frequently than others and in noun phrases DET + N and DET + ADJ + N were used. Certainly, this finding depends to a great extent on the setting of the corpus and the tasks assigned to the subjects. This finding means, however, under specific circumstances, that speakers do set linguistic preferences with respect to structural constructions. Moreover, the location of interruption often takes the position after determiners, which might have to do with the importance of definite articles in German in determining the internal grammatical coherence within prepositional and noun phrases. According to the results achieved in the corpus analysis, the linguistic category of the repair onset is often in accordance with that of the reparandum onset. The term *corrected phrase* is therefore used to represent the second part of speech repairs, indicating that the correction part is often a complete phrase, instead of corrections of single words. *Problem phrase* stands for the reparandum, also indicating that the reparandum is often a phrasal fragment which will be later corrected in terms of a newly begun phrase. In summary, this syntactic analysis clearly shows the syntax-oriented production of complex speech disfluencies. More specifically, the production is phrase-oriented.

Chapter 4

Acoustic-Prosodic Cues of Selected Types of Disfluencies

This chapter aims to investigate additional prosodic cues of disfluencies by means of concrete acoustic measurements and prosodic annotations. It has been shown in Chapter 3 that there are specific patterns of disfluencies that speakers more or less preferred to produce. As prosody plays an important role in spoken language, research work has been carried out by different means of investigation methods. All are concerned with the question: To what extent prosody is used to compensate, to override or simply to emphasize the syntactic peculiarity of utterances. Price *et al.* [93] for instance have found support for the function of prosody in disambiguating two competing syntactic structures. Large syntactic boundaries (clause boundaries) often coincide with the boundaries of major prosodic constituents. Durational information has also proved to be useful in disambiguating syntactically ambiguous constructions as mentioned in [65]. Lehiste *et al.* have used recorded sentences read by professional radio announcers as material for the perceptual experiments. However, it may be more practical to investigate what/how speakers do in "unusual" situations, namely hesitating, making errors and repairing their speech. These phenomena clearly reflect functions and characteristics which are contained in specific languages.

Investigate how disfluencies are produced with respect to prosody of disfluencies can improve understanding of the relationship between the structural (usually syntactic) and the prosodic features of specific phenomena of speech production. Although a regularity regarding the structural types of disfluencies was found in Chapter 3, the production of disfluencies depends, however, on individual choices and communication conditions available to the speakers. So, only disfluencies of specifically selected structural types produced by a single speaker are investigated in this chapter. That is to say that an acoustic-prosodic analysis will be carried out to investigate disfluencies of the forms: (*problem phrase*, *editing phase*, *corrected phrase*) and (*problem phrase*, _, *corrected phrase*). Speech repairs and repetitions with or without an editing phase will be taken into account. When the problem phrase and the corrected phrase are identical in a disfluency, it is a repetition. The reason why these two types of disfluencies are chosen is that they cover more than 80 % of the disfluencies found in the corpus.

In the first section, prosodic features and their acoustic correlates which will be empirically examined in the acoustic-prosodic analysis are introduced. Acoustic measurements and prosodic annotations are used to explore the prosodic cues provided by disfluencies. The second section deals with the details of the data used in the acoustic study. How the data are processed to be used for the acoustic-prosodic analysis is presented next, including the criteria for segmenting utterances in the selected dialogue. After the introduction on the data processing, tools used to do the acoustic measurements, as well as the annotation system for the prosodic labelling are also introduced. Statistical settings are given, because the data of measurements obtained are to be analysed by means of statistical evaluation. Results of the acoustic-prosodic study are presented in terms of three main groups: 1) F_0 features, 2) temporal features and 3) tonal phenomena. After the results have been presented, the relationship between the specific structure of selected disfluencies and their related prosodic characteristics will be discussed.

4.1 Relevant Prosodic Properties

In order to account for prosodic phenomena of disfluencies related to the lexical and higher levels, lexical and utterance prosodic features were chosen to be investigated in the acoustic-prosodic analysis. Lexical prosody is concerned with prosodic properties within or around lexical items in disfluencies, whereas utterance prosody takes into account prosodic phenomena in a wider linguistic range, for instance phrases, disfluency-phrases or utterances. Accents and duration of relevant lexical items in disfluencies are examined in order to investigate the lexical prosodic behaviour within disfluencies. With respect to utterance prosody, tonal features of the problem and the corrected phrases are compared to see whether there exists a similar or a contrastive relationship between these two parts. Furthermore, the duration of pauses between the problem and the corrected phrases is to be examined in relation to the construction type of disfluencies. The issue whether speaking tempo can be correlated with the number of troublesome words in the context of disfluencies is investigated as well. These selected prosodic features which are to be examined in the acoustic-prosodic analysis are clarified in detail in this section.

4.1.1 Prosodic Features

Two lexical prosodic features, *accents* and *item duration*, are examined. Accounting for their prosodic realization, how disfluencies are produced locally, which means what happens to the problem items and the corresponding corrected lexical items within disfluencies, will be investigated in terms of these two lexical prosodic features. A similarly produced pattern of accents of lexical entries in the same syntactic position within disfluencies provides evidence for the existence of short-time memory of prosodic pattern while speech is being processed. Information about word accents is particularly useful, when repetitions and repairs of phonological and lexical speech errors are considered. Whether the accents of the same lexical item in the problem and the corrected phrases within disfluencies remain the same, gives clues to the discussion on whether speech monitoring which is directly related to prosodic control has taken place [111].

Lehiste [64] and Lehiste *et al.* [65] have pointed out the function of specific temporal information, namely the durational features, with respect to the phenomenon of structuring syntactically ambiguous strings. Speech disfluencies are syntactically ill-formed sequences which share similar characteristics in spoken language, appearing as syntactically ambiguous constructions. Thus, it is highly possible that one can find durational information while analyzing the duration of the elements within disfluencies and therefore obtain indications of what kinds of factors are relevant to the internal structure of disfluencies. Temporal information such as duration of lexical items within disfluencies is indicative of how long the speaker needs to detect and correct an error. Temporal features often play an important role in spoken language. In this case the duration of disfluency items can provide empirical support for or against speech production models. If the items in the corrected phrase are realized in a shorter duration than the corresponding part in the problem phrase, it is possible that the monitoring has taken place earlier during the speech flow. However, it is not the main task of this thesis to prove or disprove speech production models. Results obtained by the acoustic-prosodic analysis should merely deliver preliminary evidence to whether specific phenomena in relation to the temporal features of disfluencies exist.

Three utterance prosodic features are selected to be investigated: *tonal features*, *pauses between the problem and the corrected phrases* and *speaking tempo*. Utterance prosodic features are prosodic features concerning the prosodic impression and realization of utterances. The concept "intonation" is regarded as one of the tonal features in this chapter, as well as pitch contours of smaller linguistic units. The combination of tones characterizes the details of utterance intonation. Boundary tones give information about how a speaker ends or interrupts his/her speech when disfluencies occur. This peculiarity can be investigated together with the syntactic position of disfluencies to see how they are related to each other.

Investigation of duration and location of pauses between the problem and the corrected phrases in utterances can clarify to what extent syntactic and prosodic means are used when disfluencies are made in spoken language. Furthermore, pausal duration between the problem and the corrected phrases indicates how interruptions caused by disfluencies are associated with disfluency types, whereas the length of pauses can show how long the speaker hesitates before he/she initiates his/her problematic speech.

The relationship between speaking tempo and the frequency of disfluencies produced by speakers is to be examined in order to obtain more evidence for or against the hypothesis that when speakers speak faster more mistakes are made. This phenomenon can be clearly observed by means of disfluency research.

4.1.2 Acoustic Correlates

This section deals with practical acoustic measurement. In other words, how the prosodic features mentioned above are to be acoustically measured and presented. Prosodic features and their corresponding acoustic parameters which will be measured in different positions within utterances are illustrated as follows:

The phenomenon *accent* is to be acoustically presented by *peak* F_0 which is the highest F_0 value of the accented vowel within a lexical entry. Since the surrounding prosodic environment also influences the perception of accents, *onset* F_0 *values* of relevant lexical items too are quantitatively measured. The *onset* F_0 *value* is the pitch height of the beginning of a lexical entry. The *temporal length* from the beginning to the end of lexical items is measured to account for *item duration.*

prosodic property	**acoustic measurement**
accent	peak F_0 values
	onset F_0 values
item duration	temporal length
tonal features	F_0 contour
pausal duration	temporal length
speaking tempo	words/sec

Whether the perceptual impression of *intonation* exactly corresponds to the acoustic F_0 contour is an open question, but that F_0 contour is a significant and important indicator for the representation of intonation is indisputable. *Pitch contour* within utterances containing disfluencies are qualitatively described and annotated in terms of the system *German ToBI.* German ToBI convention is used to prosodically annotate spoken utterances, which will be introduced in more detail later in this chapter. *Duration of pauses* between problem and corrected phrases are measured where available. *Speaking tempo* is calculated in terms of the quotient of the number of words in an utterance and the duration of the utterance. It should be noted that interjections in this case are counted as lexical words in the calculation.

4.2 Data and Outline of the Acoustic-Prosodic Analysis

Because the dialogues in the BAUFIX corpus and the number of disfluencies found within the dialogues are of different lengths, one female instructor from dialogue 05, was selected to be investigated. Another reason for this decision is that this rather long dialogue, about 19 minutes, provides various types of disfluencies and the instructor has spoken clearly, so that neat measurements concerning F_0 values were obtained for most of the disfluencies produced. 11

out of 37 speech disfluencies produced by the female instructor are simply interrupted without repairs. Among them, 26 are immediate speech repairs, located in different linguistic units. Some of them were repaired within one word, some within one phrase, while other repaired speech disfluencies had a length of more than one phrase. 10 single word repetitions as well were realized by the female instructor. Because this acoustic-prosodic analysis aims to examine the more complex speech repairs, i.e. repairs containing completely or partially repeated elements, speech repairs produced within one phrase are therefore not considered, as no repeated elements can be found. Furthermore, as some of the production of disfluencies lack good acoustic quality, only those 26 speech repairs and six single word repetitions produced by the instructor are taken into consideration in the acoustic analysis. These are listed below. Their translations can be found in Appendix D.

speech repairs

```
(eine gel eine gelbe schraube)
(eine auch eine rote)
(mit dem :p: mit dem)
(mit den drei :p: mit den f\"unf l\"ochern)
(mit dem :p: mit dem uh mittleren loch)
(ein ei)
(mit einmal :p: uhm mit der orangenen achkantschraube)
(in das dri in das :p: zweite)
(auf auf das teil)
(die :p: das)
(von der :p: von :p: dem)
(in in die)
(mit der :p: mit den orangenen schrauben)
(die die runde)
(mit einer :p: mit einer mutter)
(mit diesen :p: mit dem einen)
```

```
(die diese uh pl\"att)
(legst den da :p: stellst den da unter)
(die schrau die rote schraube)
(so dass es :p: so dass es)
(der :p: der fehler)
(das war uh :p: das war)
(die die lila)
(auf die :p: auf den untersten w\"urfel)
(sch uh schrauben)
(sch :p: schraubst)
```

single word repetitions

```
(was was)
(was :p: uh was)
(da da)
(einen :p: einen)
(der :p: der)
(da :p: da)
```

The transcripts which are available for the corpus are organized and presented in turns. However, *turn* is too large a unit for investigating the utterance prosody in this analysis. Therefore, utterance segmentation principles which are applied to divide spoken discourse into smaller discourse units will be introduced. Segmentation principles suggested by Nakajima & Allen [81] are adopted with some adjustments due to the specific features of the task-oriented conversations used for the analysis in this thesis. Software ESPS is used to carry out the acoustic measurements and German ToBI is applied to annotate disfluencies, especially the boundary tones.

As the surface structure of single word repetitions are different from that of speech repairs, they will be dealt with separately. Results of the analysis are to be presented in six subgroups: peak F_0 values, onset F_0 values and duration of relevant lexical items, pauses as well as speaking tempo and tonal phenomena of utterances. The aim of the first three analyses is to illustrate the relationship of the prosodic features of elements, which appear in the same or comparable phrasal position of the problem and the corrected phrases with respect to pitch height and duration by statistical means. Therefore, a paired t-test is used to check the significance of existence of an either $>$ or $<$ relation. Within t-tests, p-values have been calculated in each sub-examination for the groups of the first and second repeated words, as well as for the group of single word repetitions. Significance value is set to be $\alpha = .05$.

4.3 Utterance Segmentation

Because both lexical and utterance prosodic features are to be investigated in the acoustic-prosodic analysis, it is important that the discourse i.e. the dialogue be divided into smaller discourse units, other than turns. This kind of segmentation should be discourse-specific, because in different communication situations, different sorts of utterances are produced. It has to do with turn-taking among the interlocutors, the length of utterances and the use of pauses. The concept of *sentences* for instance is not suitable for spoken language, because utterances in spoken language often do not satisfy the grammatical rules used for written texts. As usual, stress placements which have both pragmatic and communicative functions in conversational situations are seldom found in the conventional concept of a *sentence*. Yet, these non-syntactic factors are of great importance when the description of the discourse structure of spoken language is considered. Furthermore, the discourse structure of spoken language can be more clearly presented by means of phonological and conversational structuring such as intonational boundaries, boundary tones and turn-taking than grammar. In recent work on spoken discourse structure and dialogue systems of spoken language, the concept of *utterances* are more often used than *sentences* [85]; [81]. Although there are no strong objections against the adoption of utterance units, the definition of utterance units remains yet very different [122].

Due to the difference between spoken language and written texts, I will use the term *utterance unit* in this thesis to represent the smallest discourse units containing coherent meanings applying given segmentation principles of utterance units. These principles are based on the approach suggested by Nakajima & Allen [81] and Gross *et al.* [36], i.e. some specific rules have been adjusted to better represent the type of dialogue in the BAUFIX corpus. These two approaches will first be summarized.

Gross *et al.* [36] drops syntactic and speech act features and applies merely the prosodic clues for segmentation. Signals for an utterance unit boundary are as follows:
1) a boundary tone,
2) a pause in speech longer than a single beat,
3) a resetting of the pitch level and
4) starting a new intonational phrase.

Nakajima & Allen [81] use multiple principles to segment utterance units: grammatical, pragmatic, conversational and prosodic. Grammatical principles place utterance boundaries before sentence conjunctions, whereas pragmatic principles follow speech act organization and intention of speakers. Applying conversational principles, turn-taking marks the boundary of an utterance, whereas longer pauses also separate utterances applying prosodic principles.

The contents of the dialogues in the BAUFIX corpus are segmented into utterances applying the utterance unit segmentation rules listed in Table 4.1. They account for syntactic information, since certain conjunctions are very often markers for utterance boundaries such as *and*, *or*, *so* or *but* which help identify utterances to a large extent.The BAUFIX corpus contains task-oriented conversations between two people working on building a toy plane:

one gives instructions and the other constructs the plane. New instruction units and construction parts also mark the beginning of utterances from the pragmatic viewpoint. Prosodic markers for utterance boundaries can be extended to a great variety of features such as those suggested by Gross *et al.* [36]. For reasons that some acoustic features are too vague to measure, only the feature, *long pauses*, is considered as an utterance boundary marker. Nakajima & Allen [81] also use conversational principles to segment utterance units in discourse, which is not appropriate in the BAUFIX corpus. Because instructions are often immediately completed after the interruption of the constructor, usually by brief remarks or confirmations, it does not make sense to separate this kind of discourse fragments into several parts which do not own complete meaning with respect to discourse coherence.

Table 4.1: Utterance Segmentation in the BAUFIX Dialogue

Categories	**Principles**
syntactic	An utterance ends before conjunctions such as *and*, *but* and *or*.
	An utterance ends before adverbs such as *then*, *so*.
pragmatic	An utterance ends, when an instruction unit is ended.
	An utterance begins, when a new part of the toy plane is introduced.
prosodic	An utterance ends before long pauses.

4.4 Measurements and Labelling

In order to describe the prosodic phenomena of disfluencies, the data must be prosodically annotated. This section deals with signal processing and presentation programmes used to process the utterances containing disfluencies, as well as the prosodic transcription system adopted for the prosodic labelling.

4.4.1 Tools

The programme used for the acoustic-prosodic analysis is the *Entropic Signal Processing System (ESPS)*. Disfluencies recorded in the DAT (digital auditory tape) format were identified by the author. They were then transformed into ESPS file format *.sd* files. F_0 values (*.out* files) of the digitally recorded utterances were calculated using *ESPS* tools. Their wave-forms and F_0 contours were then presented by the graphic presentation programme *xwaves*. In *xwaves*, prosodic labelling was carried out in the labelling environment *xlabel*. Figure 4.1 is an example.

4.4.2 Prosodic Annotation

There are several prosodic transcription systems available for German, for instance Selting [107], Kohler [54], Adriaens [2] and German ToBI [95]. Notations developed in these different systems are strongly related to the aim the systems are supposed to achieve; it can be a prosodic description for communication analysis, a computer readable notation system of prosody or a perception-based modelling of prosody. Accounting for the aim of the acoustic-prosodic analysis, a prosodic transcription system is needed to describe the tonal phenomena in terms of simple symbols and to illustrate the prosodic features which can not be quantitatively measured. As German ToBI provides a systematic symbolized notation system, but not in such detail as those used in communication analysis as in [107], it is chosen to annotate the data.

German ToBI is based on the ToBI (Tones and Break Indices) system which has been triggered by Pierrehumbert [89] in her PhD thesis and has become more or less the standard terminology for prosodic labelling of American English. The VERBMOBIL project has developed a ToBI-oriented prosodic system, especially for German prosody [95]; [54]. Three types of boundaries are marked: intonational, minor and irregular. Accents and tones within and

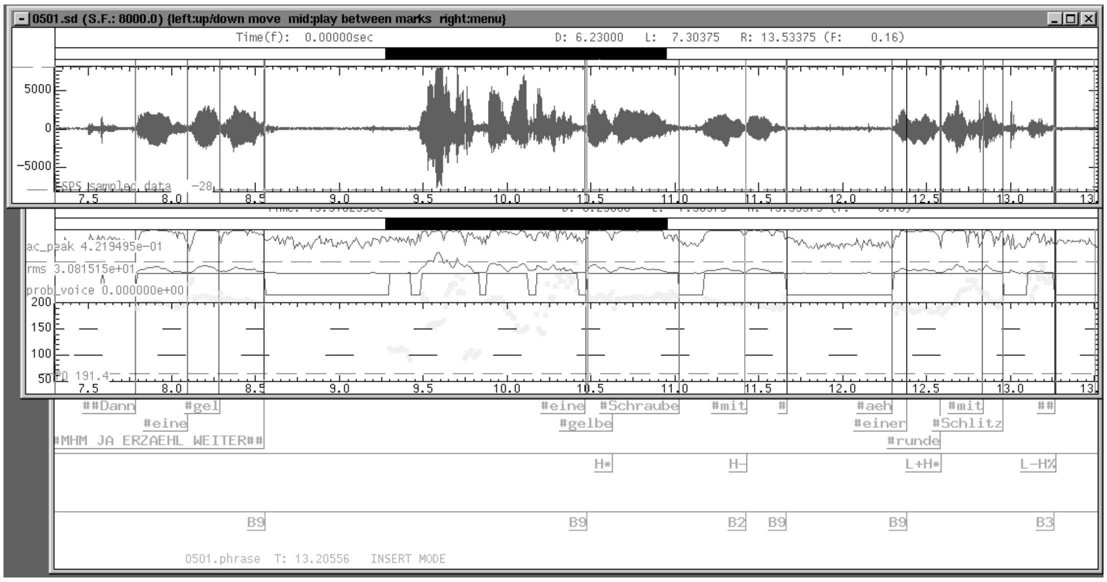

Figure 4.1: An Orthographically and Prosodically Annotated Utteranc

around boundaries are described in terms of combination of variations of high (H) and low (L) tones, for instance L+H* stands for a raised peak accent and L-L% for a terminal falling tone. The symbols and the descriptions of what the symbols stand for are presented in Table 4.2.

4.5 F_0 Features within Disfluencies

This section deals with the results related to the repeated items in the corrected phrase with respect to the peak and onset F_0 values. F_0 values around interruptions are also examined to see whether the onset F_0 value in the corrected phrase correlates with the onset F_0 value of the problem phrase or the offset F_0 value of the problem phrase.

4.5.1 Peak and Onset F_0 Values of Repeated Elements

In order to begin with the results of the acoustic-prosodic analysis, some symbols must be clarified first. As shown in Figure 4.2, the first items in the problem phrase (RP) which are repeated later in the corrected phrase (CP) are marked as I_1, whereas the corresponding items in the corrected phrase are presented as R_1. So the pair (I_1, R_1) in the speech repair, *(mit dem mit dem mittleren loch)*, is (mit, mit). Similarly, the pair (I_2, R_2) stand for the second repeated items in the corrected phrase. For instance in the repair above, (dem, dem) is the instantiated value for (I_2, R_2). That means for single word repetitions, the problem phrase is identical with the corrected phrase. (D_1, D_2) is the term used for pairs consisting of one element and the repeat of the element, e.g. *(das das)* in Figure 4.2.

Concerning the peak and onset F_0 values of the pairs (I_1, R_1), (I_2, R_2) and (D_1, D_2), no significant "greater" or "smaller" relation in the 2-tail t-test could be found with respect to the prosodic features between the first

Table 4.2: German ToBI (Reyelt/Batliner 1994)

Symbol	**Description**
B3	intonational boundary
B2	minor boundary
B9	irregular boundary
PA	primary accent
NA	secondary accent
EA	emphasized accent
H*	normal peak accent
L+H*	raised peak accent
L*+H	delayed peak accent
L*	normal low accent
H+!H*	early peak
H-	high tone
L-	low tone
L-L%	terminal fall
H-H%	question/continuation rise
L-H%	pitch rise to mid or high
H-L%	continuation fall

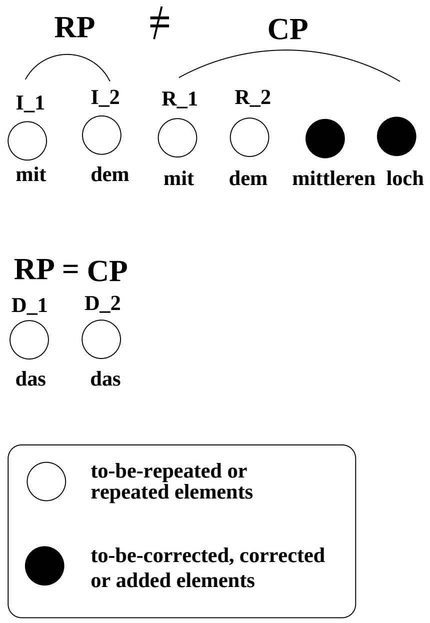

Figure 4.2: Repeated Elements in the Corrected Phrase

and second repeated items in the problem and the corrected phrases. These results are compatible with the reset hypothesis, that certain local prosodic features of the items in the corrected phrase are reset to that of the previous items in the problem phrase.

Table 4.3: Peak F_0 Values, $\alpha = .05$

	number of items	average peak F_0 (Hz) of the first element	average peak F_0 (Hz) of the repeated element	p-value
repairs				
(I_1, R_1)	24	191.25	183	$.21 > \alpha$
(I_2, R_2)	13	203.5	187.8	$.074 > \alpha$
repetitions				
(D_1, D_2)	6	191	187.2	$.269 > \alpha$

Figure 4.3 and 4.4 illustrate the distribution of the two F_0 parameters in

Table 4.4: Onset F_0 Values, $\alpha = .05$

	number of items	average onset F_0 (Hz) of the first element	average onset F_0 (Hz) of the repeated element	p-value
repairs				
(I_1, R_1)	26	191	187	$.648 > \alpha$
(I_2, R_2)	13	206.9	188.8	$.144 > \alpha$
repetitions				
(D_1, D_2)	6	195.8	200.5	$.639 > \alpha$

(I_1, R_1), (I_2, R_2) and (D_1, D_2) in terms of linear lines with the initial point having values of I_1, I_2 and D_1, respectively and the final point having values of R_1, R_2 and D_2. As shown in Figure 4.3 and 4.4, there is, on the one hand, no obvious falling/rising tendency that can be observed within the pairs. On the other hand, the values are located in a centred region in almost all the cases, with the exception of some extreme cases which are possibly due to a particular speaking intention of the speaker.

4.5.2 F_0 Values around the Interruption

Whether the prosodic features of the onset of the corrected phrases after interruptions are more similar to those of the onset of the problem phrases or rather the offset before interruptions is investigated in this section. F_0 values of the onset of the corrected phrases are compared with those of the onset and the offset of the problem phrases. Table 4.5 lists the three groups of F_0 values with their respective averages.

1-tail t-test is applied to the two groups of data containing the pitch height differences between 1) repair onset and error onset and 2) repair onset and error offset. The t-test results show no significant difference among repair onset, error onset and error offset, because p=.201 is larger than the

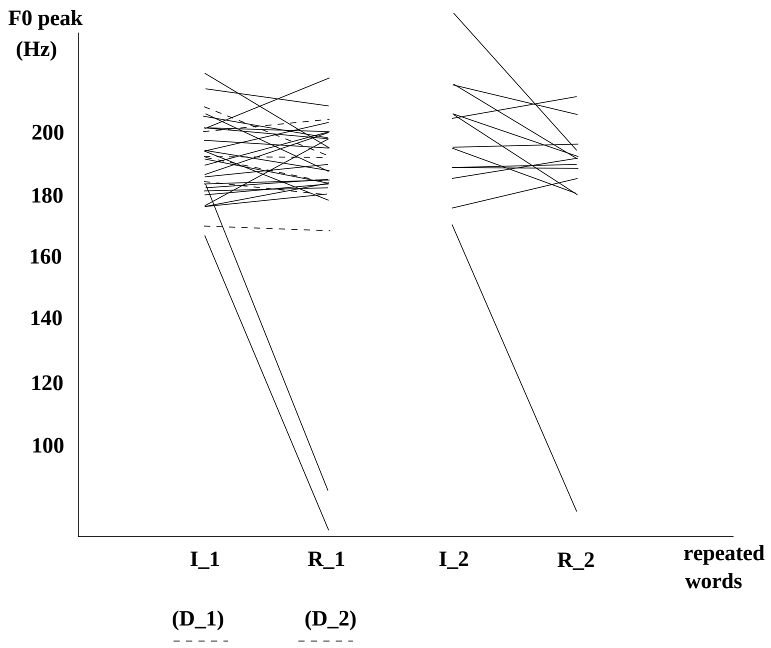

Figure 4.3: Peak F_0 Values of Repeated Elements

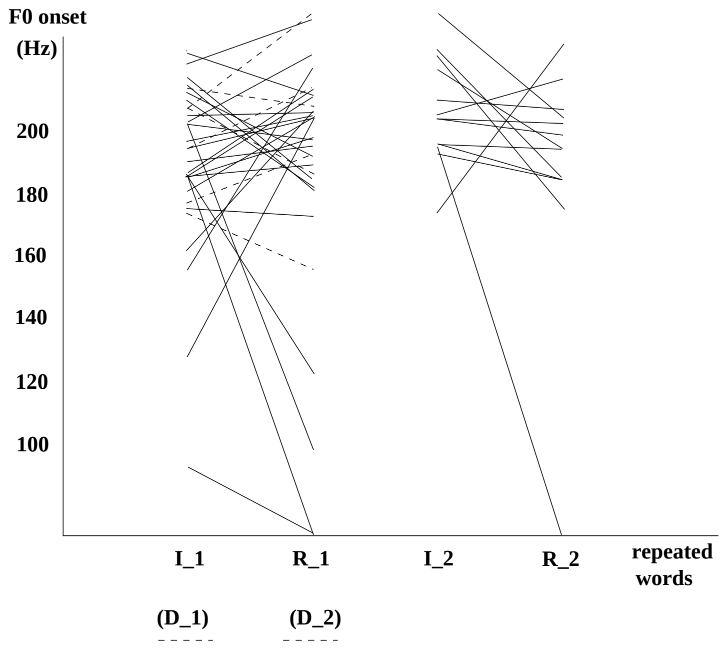

Figure 4.4: Onset F_0 Values of Repeated Elements

Table 4.5: F_0 Values around the Interruption

speech repairs	the onset of problem phrase (error onset)	the onset of corrected phrase (repair onset)	the offset of problem phrase (error offset)
1	201	197	210
2	210	182	206
3	227	213	209
4	201	99	185
5	184	190	172
6	154	220	185
7	215	181	203
8	190	196	184
9	129	204	286
10	270	263	200
11	184	211	175
12	186	122	196
13	221	243	192
14	203	205	205
15	218	185	169
16	186	67	184
17	93	69	73
18	180	204	86
19	186	212	217
20	185	198	191
21	194	203	178
22	197	203	186
23	203	227	104
24	212	191	196
25	161	205	196
26	175	172	173
mean	191	187	183

significance level α=.05. Therefore, no conclusions can be made about the question whether the pitch height of a repair onset is oriented more towards the error onset or if it is rather a continuation after the interruption.

4.6 Temporal Features

This section deals with analyses concerning temporal features of disfluencies: duration of repeated items, speaking tempo and pauses (location and length).

4.6.1 Duration of Repeated Items

Considering duration of the repeated items, the results of 1-tail t-test presented in Table 4.6 show that there exists a highly significant relation in the group of first repeated words. That is, the duration of the first elements is longer than that of the repeated elements. This implies that the speaker realized the repeated elements in a faster speaking speed. The elements therefore are clearly shortened in the corrected phrases. As Figure 4.5 illustrates, the slope of the lines drawn by the values of duration of the original and repeated elements falls in the majority of the cases. This result corresponds to that obtained by Shriberg [110]. But in the other two cases, namely the second repeated elements and the direct word repetitions, no significant empirical support for the shortened repeated elements can be found.

4.6.2 Speaking Tempo

Utterances containing disfluencies, in this case only speech repairs, produced by the female speaker are investigated with respect to the relationship between the frequency of disfluencies and the speaking tempo. In the in first column of Table 4.7, the segmented utterances are numbered according to the signal files in which they were recorded. The number of words in

Table 4.6: Duration of Items, $\alpha = .05$

	number of items	average duration (sec) of the first element	average duration (sec) of the repeated element	p-value
repairs				
(I_1, R_1)	24	.235	.175	$.009^* < \alpha$
(I_2, R_2)	12	.244	.203	$.123 > \alpha$
repetitions				
(D_1, D_2)	6	.246	.213	$.288 > \alpha$

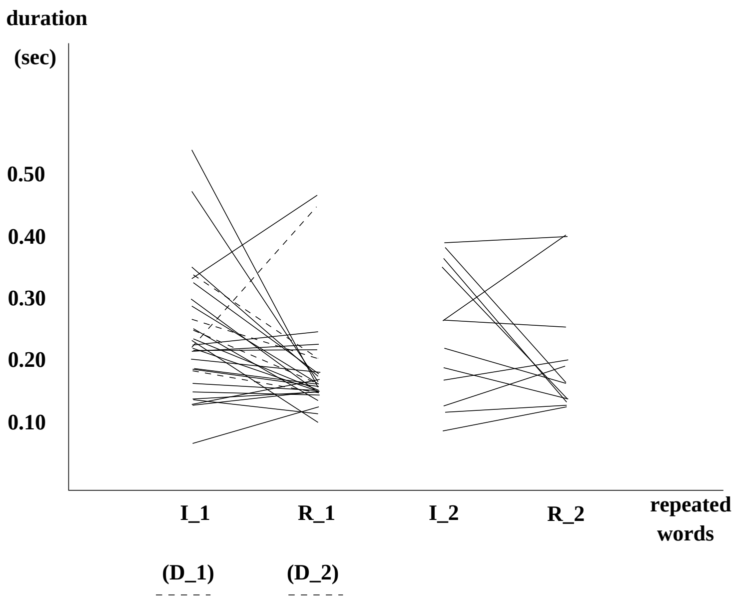

Figure 4.5: Duration of Repeated Elements

disfluencies and utterances as well as the proportion of disfluency words in an utterance are listed in the following columns. Speaking tempo is presented as the average number of words per second and is also included in the table[1].

The quotient of the number of words in each disfluency and the number of words in the corresponding utterance is compared with the speaking tempo, the result is illustrated in Figure 4.6. There is no correlation found (correlation coefficient r= -0.148) between the speaking tempo of utterances and the frequency of disfluencies within utterances. As the distribution of coordinates (speaking tempo, disfluency) shows neither a *positive* nor a *negative* correlation The hypothesis that when one speaks faster, more disfluencies are produced is not empirically supported.

4.6.3 Pauses

Duration and location of filled and unfilled pauses within complex speech disfluencies (speech repairs in this case) are examined in dialogue 05. As in the previous sections, only utterances produced by the instructor are examined due to the fact that pausal properties are speaker-specific and therefore the use of pauses can vary to a large extent among different speakers. The result concerning duration and location of pauses is presented in Table 4.8, where **A** stands for location between the problem and the corrected phrases, **E** within the problem phrase and **R** within the corrected phrases.

As Table 4.8 shows, when pauses are produced within speech repairs, their most frequent location is between the problem and the corrected phrases.

[1] For some disfluencies, the speaking tempo could not be calculated, because the instructor and constructor have spoken simultaneously and the temporal boundary can not be properly set. Those cases are excluded.

Table 4.7: Speaking Rate vs. Frequency of Disfluencies

utterances	words in utterance	words in disfluency	speaking tempo (word/sec)	$\frac{\text{words in disfluency}}{\text{words in utterance}}$
0501-1	9	4	1.73	.44
0501-2	12	11	3.36	.92
0501-3	27	7	3.45	.26
0503	26	12	3.38	.46
0504	14	2	3.76	.14
0505	23	7	2.67	.30
0507	10	3	3.17	.30
0508-1	9	1	5.24	.11
0508-2	25	7	2.42	.28
0509	16	4	3.19	.25
0510	8	2	5.03	.25
0511	15	4	4.17	.27
0513-1	14	2	3.57	.14
0513-2	25	4	2.98	.16
0514	10	3	4.29	.30
0515	10	5	3.46	.50
0516	27	4	3.75	.15
0517	10	6	3.8	.60
0518	25	6	3.76	.24
0519	19	10	4.61	.53
0522	10	4	4.17	.40
0523	15	3	2.44	.20
0525	17	5	4.03	.29
0526	9	4	4.02	.44
0527	26	6	4.59	.23
0528	6	3	4.39	.50
0529	26	2	4.13	.08
0533	11	2	3.8	.18
0534	18	4	4.39	.22
0535-3	12	7	4.22	.58
0536-1	17	10	3.23	.59
0536-2	14	2	5.08	.14
0537	6	2	4.41	.33

Table 4.8: Duration and Location of Pauses

utterances	pauses	duration (sec)	location
	filled pauses		
0501-2	äh	.09	A
0505	uh	.331	R
0508-2	uhm	.325	A
0518	uh	.574	A
0522	uh	.309	R
0531	uh	.223	A
0536-1	uh	.409	E
	unfilled pauses		
0501-2		.628	A
0501-3		0.571	A
0503		.437	A
0503		.108	A
0505		.227	A
0508-2		1.053	A
0509		.668	R
0513-1		.06	A
0513-2		.205	A
0413-2		.119	R
0517		.288	A
0518		.208	A
0519		.297	A
0519		.248	A
0524		.246	A
0526		.223	E
0527		.174	A
0528		.187	A
0531		.166	A
0534		.125	A
0536-1		.424	A
0536-2		.269	E

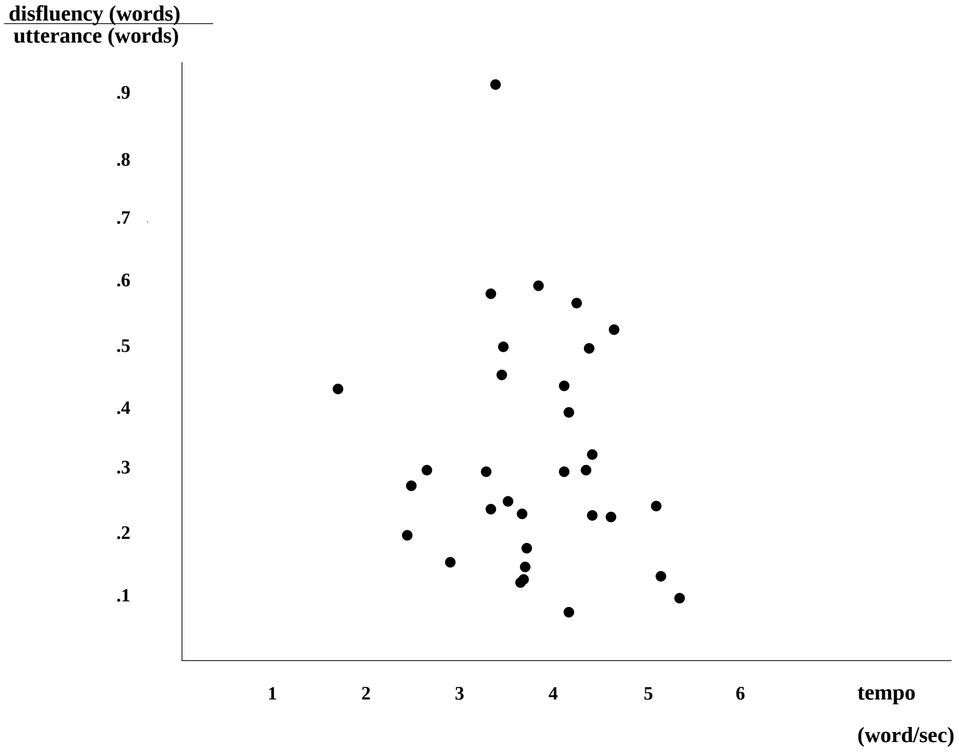

Figure 4.6: Speaking Tempo vs. Disfluencies

This result supports the proposal that the unit *phrase* is an adequate unit for describing the structure of disfluencies. Pauses are seldom used within the problem or corrected phrases in speech repairs. And the results also show that pauses located between phrases are not necessarily longer than those within the problem or corrected phrases. This indicates that the use of pauses can vary to a great extent in different speaking situations with different speaker intentions and that it is possibly "dangerous" to use pauses in spoken language as an indicator for utterance boundaries without further specifications.

4.7 Tonal Phenomena

Tonal patterns in the problem and the corrected phrases within selected speech repairs and repetitions are compared to see whether there exist similar or contrastive melodic patterns in these phrases. Not all the disfluencies which have been investigated in the previous sections are chosen to be examined in this analysis, for the reason being that some of them can not provide relevant information about tonal patterns. For instance, speech repairs with no repeated elements do not provide any clues to the difference or similarity with respect to tonal phenomena within disfluencies.

4.7.1 Similar and Contrastive Tone Patterns

Selected speech repairs and repetitions including their tonal features are shown in Figure 4.7. The correlation between similar/contrastive tones and speech repairs/repetitions is then statistically tested. The results are presented in Table 4.9. Statistically, the correlation between similar/contrastive tones and repeated/repaired elements is significant according to the fisher exact test (1-tail), with $p=.0387* < \alpha=.05$. The results show the tendency that when words are simply repeated, similar tone pattern is more frequently used, whereas a contrastive tone is produced more often in the case of overt repairs.

Table 4.9: Correlation of Tones and Disfluency Types

	repeated	**repaired**	total
similar tones	9	2	11
contrastive tones	2	5	7
total	11	7	18

problem phrase	**corrected phrase**
[mit,dem] **L-**	[mit,dem] **L-**
[mit,den,drei] **L-**	[mit,den,fuenf,loechern] **H-**
[mit,dem] **L-**	[mit,dem,uh,mittleren,loch] **L-**
[in,das,dri] **L-**	[in,das,zweite] **H-**
[von,der] **H-**	[von,dem] **L-**
[mit,der] **L-**	[mit,den,orangenen,schrauben] **L-**
[mit,einer] **L-**	[mit,einer,mutter] **H-**
[mit,diesen] **H-**	[mit,dem,einen] **L-**
[legst,den,da] **L-**	[stellst,den,da,unter] **H-**
[so,dass,es] **H-**	[so,dass,es] **H-**
[das,war] **H-**	[das,war] **H-**
[auf,die] **H-**	[auf,den,untersten,wuerfel] **H-**
[was] **L-**	[was] **H-**
[was] **H-**	[was] **H-**
[da] **H-**	[da] **H-**
[einen] **L-**	[einen] **L-**
[der] **H-**	[der] **H-**
[da] **H-**	[da] **H-**

Figure 4.7: Tonal Patterns of Repairs and Repetitions

4.7.2 Illustrations

To make clear the difference between similar and contrastive tones, one utterance is selected to illustrate this point. The recorded utterance in 0503 is *uhmm jetzt fängst du mit dem mit dem an mit dem drei mit den fünf löchern*[2]. And this utterance contains a direct repetition *mit dem mit dem*[3] and a speech repair *mit den drei mit den fünf löchern*[4]. As illustrate in Figure 4.8, the tonal patterns of both parts in the repetition are similar. On the contrary, the speech repair *mit dem drei mit dem fünf löchern* is realized in the way that the speech error *drei* is accompanied by a low tone and the repair *fünf* is emphasized by a high tone.

Another result concerning the relationship between semantic and prosodic properties in disfluencies were mentioned by Levelt & Cutler [69] that erroneous utterances are more frequently prosodically marked than simply inappropriate terms. The result achieved here is based on the F_0 pitch contours presented in *xwaves* with the auditory judgements of the author. Nevertheless, it corresponds to the perceptual examination done by Levelt & Cutler, although it emphasizes another aspect. Levelt & Cutler investigated the relationship of semantics and utterances containing erroneous/inappropriate terms, whereas in this analysis the focus is on the relationship between semantics and the tonal patterns of the repeated/corrected items. What can be concluded regarding these two results is that the semantic contents play a role in the prosodic realization of disfluencies in spoken language.

[2]uhmm now you begin with that with that with three with five holes
[3]with that with that
[4]with the three with the five holes

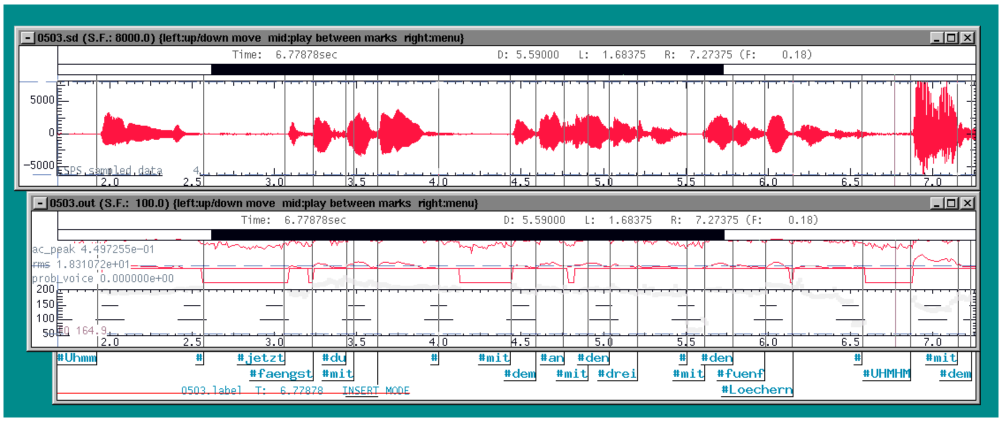

Figure 4.8: Similar and Contrastive Tone Patterns

4.8 Prosodic Marking within Complex Disfluencies

This section focuses on the results of the acoustic-prosodic analysis of the relationship between the prosodic marking and the structure of selected disfluencies.

4.8.1 Onset F_0 Values vs. Corrected Phrases

While restarting a phrase to correct the previous erroneous phrase, the speaker tends to re-use a similar pitch height of that realized at the beginning of the problem phrase. Although the data obtained above are in accordance with the reset hypothesis [67]; [111]; [73], the phenomenon of F_0 reset, as a null hypothesis, can not be directly proved by statistical means. The fact that no significant difference exists in F_0 values can be used to identify the existence of a corrected phrase. If the F_0 contour does not follow the expected declination of utterances, there possibly exists a complex disfluency, namely a speech repair. Shriberg [110]; [111] described a similar phenomenon related to direct repeats, where she used the F_0 restart as a cue to differentiate the perspective and retrospective repeats mentioned by [41].

4.8.2 Temporal Marking vs. Pattern of Corrected Phrases

The first repeated element in the corrected phrase is shorter than the corresponding item in the problem phrase. Why does the speaker speak more quickly after he/she interrupts the erroneous sequence and begins the correction? This can be an indication supporting the fact that a speaker processes his/her speech before he/she produces the sequence. Before the problem phrases are interrupted, the speaker in fact already knows what he/she will say next. So the speaking tempo is faster than before. This faster speaking tempo also marks the irrelevance of the information contained within the strings of the repeated elements in the corrected phrases. The reason for this is that the repeated elements in fact do not offer any new information, and the listener can therefore have quicker access to new in-coming information.

This result also helps to identify the internal structure of the corrected phrase, especially in natural language processing. The number of words repeated in the corrected phrase is related to the syntactic pattern of the corrected phrase, and forms the basic pattern of disfluencies. Similar results have been obtained by Shriberg [111], although she dealt with direct word repetitions. Shriberg's results support the notion that the repeated words are shorter then the previous ones, when direct word repetitions are considered. According to the results obtained in the acoustic-prosodic analysis in this chapter, this phenomenon of shortened elements can also be found in the context of speech repairs where retracing takes place.

4.8.3 Tonal Marking vs. Correction/Repetition

When there are no elements which need to be corrected within disfluencies, similar tonal patterns in problem and corrected phrases can be observed. Opposingly, corrected elements in the corrected phrase are frequently realized with a contrastive tonal pattern in comparison to that of erroneous elements in the problem phrase. This prosodic marking signals the importance of the process of correction in the corrected phrase, i.e. the relevance of the information. Contrastive tones serve the function of attracting the attention of listeners to emphasize that the incoming words should repair the previous erroneous speech strings, indicating that new information is being produced. This is different from the case of simple repetitions, whereby no new information is included.

4.9 Summary

In Chapter 4, a prosodic-acoustic analysis was carried out to gather more lingusitic features inherited in the production of speech disfluencies, especially prosodic features. Data produced by a female subject, acting as instructor in co-operation with a constructor, were selected to be examined. Utterances containing disfluencies were identified and segmented. In order to execute the prosodic-acoustic analysis, data were digitally recorded in ESPS format and analysed. F_0 values and durational features were measured using ESPS; data were also prosodically annotated in terms of German ToBI. The prosodic-acoustic analysis contained three main areas: F_0, temporal and tonal features of disfluencies.

As suggested in the reset hypothesis in Chapter 2, no distinctive differences in peak and onset F_0 values were found between the original words in the problem phrases and the repeated words in the corrected phrases which are at the same time the first and second words in the trouble and the corrected phrases respectively. In other words, the intonational features at the beginning of the problem phrase and the corrected phrase were not significantly different. This result implicitly supports the reset hypothesis that when speech repairs are initiated, the declination development is reset to certain earlier positions, in the sense that the global baseline of the utterance does not decline in the way it should be. Concerning the temporal features, the first repeated word in the corrected phrase was statistically significantly shorter than the original one in the problem phrase. This can be regarded as a kind of temporal marking of the "being produced again" of a phrase with repeating or repairing features. The relationship between speaking tempo and the number of speech disfluencies was also investigated, however, no significant correlation could be found. Furthermore, pauses were more often located *between* the problem and the corrected phrases, instead of *within* the problem and the corrected phrases themselves. But no significant correlation could be found between the duration of pauses and their locations.

Despite the moderate sample, a clear correlation between the occurrences of repetition vs. correction and the patterns of similar vs. contrastive tones was found. Tonal marking in this case is used to mark and emphasize the new in-coming information. The three main findings: support for the reset hypothesis, temporal marking of the existence of a re-initiation and tonal marking of contrasting repetition from correction, are prosodic factors which are directly related to the production of speech disfluencies.

Chapter 5

Re-Analysing Speaker-Dependent Syntactic Features

The corpus analysis concerning the syntactic features of disfluencies was done by investigating all 22 dialogues contained in the BAUFIX, whereas the acoustic-prosodic analysis was carried out by studying the data produced by only one selected speaker among the 44 speakers in the corpus. This thesis aims to focus on syntactic and prosodic factors of disfluencies among a large number of factors influencing speech production. Empirical evidence related to these two types of features were already presented in Chapter 3 and 4. Because the investigating data obtained were of different sizes, this chapter re-examines the syntactic features in the prosodic analysis data. The goal of this re-analysis is to make sure that the two types of information (syntactic and prosodic) provided by the analyses can be re-found over the same corpus data.

5.1 Speaker-Dependent and Speaker-Independent Results

In this syntactic re-analysis, the factors investigated in Chapter 3 are to be examined once more by restricting the data to those produced by the female subject in dialogue 05, whose utterances containing disfluencies have been examined in the acoustic-prosodic analysis. The length of disfluencies, their syntactic category, the construction types of PP and NP disfluencies, the location of interruption and the onset and the offset of problem and corrected phrases are investigated. The results will be shown in comparison to the results obtained in Chapter 3. Data produced by the instructor in dialogue 05 is abbreviated by I-05, whereas BAUFIX denotes the whole corpus. Syntactic features of speech disfluencies produced by I-05, which are speaker-dependent, are compared with those found with regard to the entire corpus data which are speaker-independent.

5.1.1 Syntactic Length and Category

Similar to the results obtained in Chapter 3, the syntactic length of disfluencies produced by I-05 is in most cases longer than one phrase, as listed in Table 5.1. The distribution of disfluency length is very similar in both groups of data, although the groups are of different sizes. 44 subjects made up the BAUFIX corpus, while I-05 consists only one single subject. Figure 5.1 illustrates this similar distribution, comparing the length of speech disfluencies. In both cases, disfluencies realized within one word or one phrase are about 25% out of the speech disfluencies produced in BAUFIX and I-05. As a result, speech disfluencies having a length of more than one phrase are also in the majority in both of the data.

As shown in Table 5.2, disfluencies produced in I-05 are more frequently located in the prepositional and noun phrases, as found in the BAUFIX

Table 5.1: Syntactic Length of Disfluencies

	BAUFIX		**I-05**	
category	**frequency**	**percentage**	**frequency**	**percentage**
I.	54	10.8%	6	10.3%
II.	84	16.8%	9	15.5%
III.	362	72.4%	43	74.2%
total	500	100%	58	100%

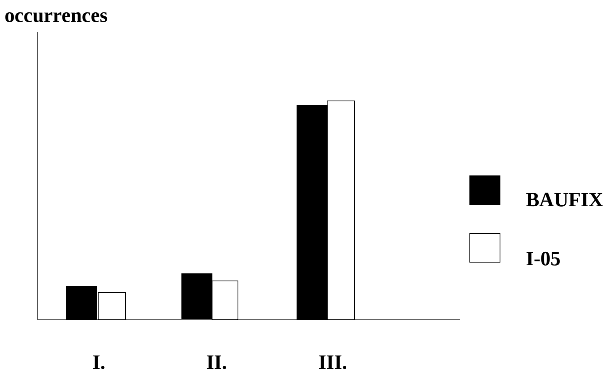

Figure 5.1: Comparison of Linguistic Length in BAUFIX and I-05

corpus. A possible reason why this phenomenon was observed is that the instructor had to describe the objects exactly in order to make her speech clear, so that the constructor could understand which objects were necessary for the construction. Thus, noun phrases and prepositional phrases show more necessity to be repaired, if they are incorrectly or inappropriately spoken. Similarly, the percentage of other categories such as verb phrases and adverbs also shows the same distribution in BAUFIX and I-05. Again, this phenomenon can be more clearly observed in a graphic presentation, as shown in Figure 5.2.

Table 5.2: Disfluencies in Syntactic Category

	BAUFIX		**I-05**	
category	**frequency**	**percentage**	**frequency**	**percentage**
PP	174	34.8%	21	36.2%
NP	190	38%	23	40%
VP	32	6.4%	3	5.1%
ADV	31	6.2%	2	3.4%
2P	38	7.6%	6	10.3%
OTHER	35	7%	3	5.1%
total	500	100%	58	100%

5.1.2 Internal Structure of PP and NP Disfluencies

With regard to the internal structure of PP and NP disfluencies, similar construction types are found between BAUFIX and I-05, as shown in Table 5.3 and 5.4. PP_2, PP_4 and PP_5 disfluencies constitute the three largest groups in I-05. Compared with the results of BAUFIX, disfluencies produced in the form of PP_2 and PP_5 (definition cf. Chapter 3) are among the more frequently produced construction types in I-05. Similarly, NP_2 and NP_4 are

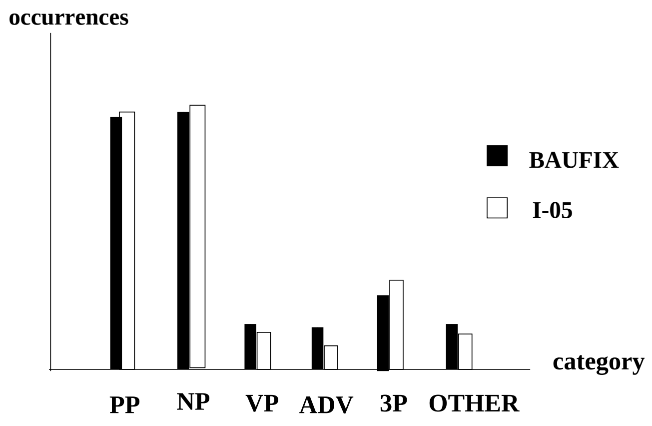

Figure 5.2: Comparison of Linguistic Categories in BAUFIX and I-05

also realized more frequently than other types of NP disfluencies in both cases. Comparisons of internal construction types of PP and NP speech disfluencies in BAUFIX and I-05 are graphically illustrated in Figure 5.3 and 5.4.

Table 5.3: Phrasal Construction in PP

	BAUFIX	**I-05**
construction type	**frequency**	**frequency**
PP_1	5(3%)	0(0%)
PP_2	46(26%)	3(14%)
PP_3	4(2%)	0(0%)
PP_4	13(8%)	3(14%)
PP_5	47(27%)	7(33%)
PP_rest	59(34%)	8(38%)
total	174	21

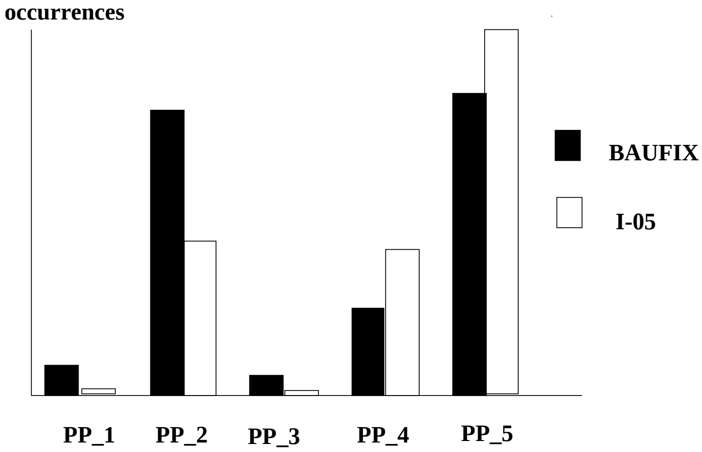

Figure 5.3: Construction Types of PP in BAUFIX and I-05

Table 5.4: Phrasal Construction in NP

	BAUFIX	**I-05**
construction type	**frequency**	**frequency**
NP_1	5(3%)	0(0%)
NP_2	51(27%)	5(22%)
NP_3	19(10%)	1(4%)
NP_4	49(26%)	5(22%)
NP_5	11(6%)	0(0%)
NP_6	12(6%)	0(0%)
NP_rest	43(23%)	12(52%)
total	190	23

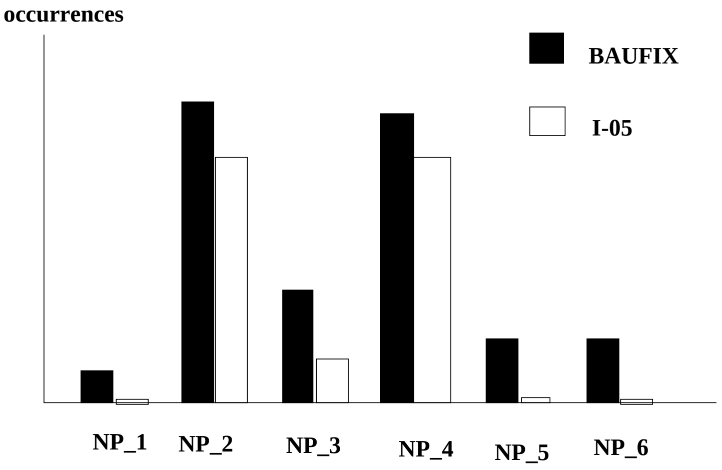

Figure 5.4: Construction Types of NP in BAUFIX and I-05

5.1.3 Location of Interruption

In PP disfluencies, nine out of 21 cases (43%) are interrupted after determiners and five after prepositions (24%), whereas in NP disfluencies 16 out of 23 cases (70%) are interrupted after determiners. To summarize the previous results of the BAUFIX corpus, 44% of the PP disfluencies are interrupted after determiners, whereas 25% are interrupted after prepositions. 63% of the NP disfluencies are interrupted after determiners. Referring to the percentage of the location of interruptions, the BAUFIX and I-05 data show highly similar distributions in this comparison.

5.1.4 Repair Onset

Throughout the BAUFIX corpus, repaired disfluencies located in PP and NP are in most cases initiated at the beginning of a phrase. Highly similar results can be observed regarding the I-05 data. In 15 out of 16 repaired PP disfluencies, the repair is initiated at the phrasal beginning positions, whereas all 19 repaired NP disfluencies are initiated at the beginning of a phrase. The comparison of the results is shown in Table 5.5.

Table 5.5: Onset of the Corrected Phrase

	BAUFIX		**I-05**	
	repaired disfluencies	**initiated at phrasal onset**	**repaired disfluencies**	**initiated at phrasal onset**
PP	120	120 (100%)	16	15 (94%)
NP	141	139 (99%)	19	19 (100%)

5.1.5 Repair Offset

While the repair onset coincides very frequently with the phrasal initial position in the BAUFIX corpus, the correspondence between the repair offset and the phrasal final position is still significant but less frequent than in the case of the repair onset. In I-05, there are 12 cases out of 16 PP repaired disfluencies in which the repair offset is realized at the phrasal final position. In the case of NP disfluencies, 10 out of 19 repaired NP disfluencies end at the phrasal final position. The distribution is similar to that of the BAUFIX corpus. The result is summarized in Table 5.6.

Table 5.6: Offset of the Corrected Phrase

	BAUFIX		**I-05**	
	repaired disfluencies	**ended at phrasal offset**	**repaired disfluencies**	**ended at phrasal offset**
PP	120	97 (81%)	16	12 (75%)
NP	141	107 (76%)	19	10 (53%)

5.2 Similar Syntactic Features in BAUFIX and I-05

All the syntactic characteristics found within the BAUFIX corpus were clearly re-found in the I-05 data. To sum up, in total five points of consistency between BAUFIX and I-05 data were found:
1) The length of disfluencies in phrases is in most cases longer than one and shorter than three. That is, a disfluency is usually composed of two phrases, a problem phrase which is usually a disfluency-phrase as defined in Chapter 3 and a corrected phrase.
2) With respect to the syntactic category of the disfluencies examined, it is more likely to find disfluencies within PP and NP constructions than in any other categories.
3) Errors and repairs related to prepositional phrases are produced more frequently in phrasal structures such as P + DET + N and P + DET + ADJ + N. Interestingly, NP disfluencies are found more often in DET + N and DET + ADJ + N, which have the same internal structure as in the prepositional phrases.
4) That interruption occurs more often after prepositions and determiners is also the case regarding the I-05 data.
5) The same features concerning repair onset and offset are found in both cases too, namely repair onset is phrasal-initial and repair offset is phrasal-final.

Results of this syntactic re-analysis convincingly show high agreement with those in Chapter 3. That is, syntactic features of disfluencies found by investigating the whole corpus in the BAUFIX corpus are the same as those in the selected material. With respect to every factor examined, the results of this re-analysis are in accordance with those of the previous syntactic analysis. Because the syntactic features found in the BAUFIX corpus data can be re-found for the chosen speaker of dialogue 05, it is therefore empirically justified to apply both the syntactic and prosodic information together to describe the structural characteristics of complex disfluencies.

However, it is important to notice that the results found in the previous analyses can only provide clues to disfluencies which were produced in the specific task-oriented conversations contained in the corpus. Statements made in this thesis are not aimed to generally explain the linguistic behaviour of speech disfluencies in spoken language, because the data used in the analyses were collected under experimental conditions. Labov [58] has given specific stylistic criteria for defining spoken language forms: casual, careful and spontaneous speech. *Spontaneous speech* defined in [58] is a kind of casual speech, but in a formal context. Recently, a number of large corpora have been collected. However, not much attention has been given to the sociolinguistic factors of the data.

The formal description and modelling of speech disfluencies carried out in the next chapter should serve as a pioneer attempt to integrate and conclude the syntactic and prosodic cues of speech disfluencies and therefore show a new point of view in the relationship between syntax and prosody in spoken language.

5.3 Relationship between Syntax and Prosody

As the simultaneous appearance of both syntactic and prosodic use has been confirmed by re-examining the syntactic features of speech disfluencies produced by I-05, this section aims to discuss briefly the role of prosody in spoken language focusing on the use of prosody in case of disfluencies. Prosody is present on every linguistic level. Lexical items carry lexical prosodic properties such as primary or secondary word stresses or lexical tones, whereas phrasal boundary tones as well as utterance stresses can be determined to segment and structure utterances. Not to mention the importance of intonation in spoken language. Under specific circumstances, prosody such as intonation patterns can be adopted to overwrite the original sentence meaning, although prosody in usual communication situations often plays the role of supporting and underscoring the syntax-based utterance structures. One can apply prosody to express particular intentions in normal utterance structure, without having to choose complicated and difficult utterances to make unusual intentions clear. However, as shown previously in the three empirical studies concerning speech disfluencies in spoken dialogues, both syntactic and prosodic characteristics were found in the production of speech disfluencies. That is, where syntax shows an effect, there is also a prosodic effect available, dealing with the production of speech disfluencies.

In the case of speech repairs, it is important to emphasize or attract the listeners' attention that something wrong has been said in the course of the speaking and it is at this given moment being repaired. A prosodic effect was found to contrast given from new information by using temporal and tonal markings. More specifically, repeated words in the corrected phrase are spoken more quickly than the previous originals, whereby word repetitions in fact don't supply any new information. In contrast to the case of repetitions, repaired words are tonally marked. Repaired words in the corrected phrase contain new information and have to distinguish themselves from the previously given information which has been wrongly or inappropriately verbalized. It is, therefore, useful in emphasizing the importance of the ongoing new information by marking the words prosodically. Prosody used on such occasions shows a syntax- and semantics-accompanying function, where prosody in no way replaces the syntactic and semantic tasks.

Based on the results of the empirical studies, the question whether prosody plays an autonomous role in spoken language can not be directly answered. One possible method is to design experiments which are often applied in psycholinguistic investigations to test different hypotheses in this field. However, in the studies carried out in this thesis up to now, one can clearly observe the supporting role played by prosody in the production of complex speech disfluencies. This function of prosody is supposed to "help" listeners differentiate important from less important information, temporarily relevant from irrelevant information and distinguish spoken language from written language where one does not have the possibility to use prosody to mark the degree of importance of information. Furthermore, a related question which cannot be directly answered in this thesis but arises in the empirical results is: to which extent can prosodic marking really help listeners understand disfluent speech and effectively pick out the right information?

5.4 Summary

In Chapter 5, the syntactic analysis carried out in Chapter 3 was repeated by using the same material used in the acoustic-prosodic analysis. This re-analysis aimed to test whether the syntactic results found in the speaker-independent corpus could also be found in the speaker-dependent data. Another purpose for carrying out this re-analysis was to make sure that the syntactic and prosodic features found previously by investigating different sizes of corpora could beapplied to the same data. The results of the re-analysis were to a great extent in accordance with those obtained in Chapter 3. Prepositional and noun phrases were the two most frequent categories in which disfluencies were produced. Regarding the internal structure of complex disfluencies, i.e. the construction types and the location of interruption, similar distributions were clearly found in both data.

Chapter 6

Modelling Disfluencies

This chapter is concerned with a preliminary proposal for describing and modelling the occurrence of simple and complex speech disfluencies in spoken dialogues, based on the empirical results obtained in previous chapters and on the related theoretical considerations dealing with spoken language. It is, however, not the goal of this thesis to develop a sophisticated model for all possible variations of disfluencies in spoken language. The description system to be suggested in this chapter should provide a systematic and in particular, syntax-oriented theoretical approach to handling the most frequent types of occurrences of disfluencies found in the BAUFIX corpus. Syntactic and prosodic cues are the main information used to describe selected types of disfluency. The reason for applying these two sorts of information is that both effects were clearly found earlier in the corpus analysis.

The description of speech disfluencies starts with specific observations in the internal construction of the most frequent types of speech disfluency found in PPs and NPs. Subsequently, by investigating these speech disfluencies found in the corpus, its aim is to look for regularity on the surface structure level and to select relevant parameters for describing the disfluencies. Disfluency-phrase (DFP) which was introduced in Chapter 3 and disfluency structure (DFS) are therefore chosen to denote the structural description of speech disfluencies, where the definition of disfluency-phrase is more specified than when it was used earlier as an operative definition. Significant features inherited in the production of complex disfluencies, such as the retracing of a repairing process, the interruption after interrupted words and the completion of speech repairs are modelled in terms of non-deterministic finite state automata [45] with ϵ-transitions. This modelling covers a large majority of occurrences found throughout the corpus, as will be shown in the final section of this chapter.

6.1 Structural Characteristics of Complex Disfluencies

Similar to the results obtained by Daly-Kelly [17] that pauses often occurred within noun and prepositional phrases, speech disfluencies identified in the corpus data were clearly produced more frequently in the prepositional and noun phrases than in any other syntactic category. This finding is particularly characteristic of the task-oriented dialogues of this corpus, in which appositive noun and prepositional phrases were frequently used to describe details of objects. On the other hand, very few verbs were needed for describing actions, because the actions which should have been undertaken were not many, or rather, instructors tended to use simple descriptions of the actions. As a result, disfluencies within or around the verb phrases are extremely seldom, compared with the prepositional and noun phrases. Although the results presented in this thesis are to a great extent corpus-dependent and the distribution of types of speech disfluency may therefore be different from corpus to corpus, the internal structure of collected speech disfluencies nevertheless provides general information about which types of speech disfluencies are preferred. Thus, in order to obtain a clearer insight into possible regularities of the production of disfluencies, we will first look at the more frequently produced disfluencies in PPs and NPs which were found in the BAUFIX corpus.

6.1.1 Complex Disfluencies in PP

As illustrated in the results of the syntactic analysis in Chapter 3, disfluencies related to prepositional phrases were divided into six groups according to their underlying phrasal structure: PP_1, PP_2, PP_3, PP_4, PP_5 and PP_rest. Among them, PP_2 and PP_5 are clearly significantly larger than the other groups. Therefore, we will first have a look at all of the occurrences of disfluencies produced in the form PP_2 and PP_5.
PP_2: PP => P + DET + N
PP_5: PP => P + DET + ADJ + N

In total, 46 occurrences were produced in the form of P DET N, including simple (interrupted problem phrases) and more complex (repaired) speech disfluencies, whereas 47 occurrences of the form P DET ADJ N were found. Some notations used to list the occurrences are to be defined first. *N-d* stands for a fragmentary noun and *ADJ-d* is an adjective which is not realized completely. P(DET) represents sequences such as *zum*[1] or *zur*[2] which are abbreviation forms for the combination of a preposition and a determiner. The results are illustrated in Table 6.1.

This result is in accordance with that presented in Chapter 3, that the majority of complex disfluencies were produced in combination of a disfluency-phrase and a corrected phrase. 39 out of 46 in the case of PP_2 and 38 out of 47 in the case of PP_5 correspond to this structure. Only 4 out of 43 repaired disfluencies in PP_2 occurred within one single phrase and 7 out of 45 repaired PP_5 disfluencies were produced in one phrase. Moreover, the maximal word length in these two PP disfluencies is 8, in which the whole

[1]zum is the abbreviation form for *zu* and *dem*, where *zu* means onto and *dem* is a determiner (singular, dative, masculine/neutral).

[2]zur is the abbreviation form for *zu* and *der*, where *zu* means onto and *der* is a determiner (singular, dative, feminine).

Table 6.1: Occurrences of PP_2 and PP_5 in BAUFIX

P + DET + N		**P + DET + ADJ + N**	
repaired		repaired	
(P, **P**, DET, N)	10	(P, **P**, DET, ADJ, N)	7
(P, **P**, DET, *N-d*)	1	(P, DET, **P**, DET, ADJ, N)	16
(P(DET), **P**, DET, N)	1	(P(DET), **P**, DET, ADJ, N)	1
(P(DET), **P**(DET), N)	1	(P(DET), **P**(DET), ADJ, N)	1
(P, DET, **P**, DET, N)	23	(P, DET, ADJ, **P**, DET, ADJ, N)	8
(P, DET, *N-d*, **P**, DET, N)	2	(P, DET, *ADJ-d*, **P**, DET, ADJ, N)	2
(P(DET), *N-d*, **P**(DET), N)	1	(P(DET), *ADJ-d*, **P**(DET), ADJ, N)	1
		(P, DET, ADJ, N, **P**, DET, ADJ, N)	2
(P, DET, DET, N)	2	(P, DET, DET, ADJ, N)	3
(P, DET, *N-d*, N)	2	(P, DET, DET, DET, ADJ, N)	1
		(P(DET), *ADJ-d*, ADJ, N)	1
		(P, DET, ADJ, *N-d*, N)	1
		(P, DET, *N-d*, ADJ, N)	1
only interrupted		only interrupted	
(P, DET, *N-d*)	2	(P, DET, ADJ, *N-d*)	1
(P(DET), *N-d*)	1	(P(DET), ADJ, *N-d*)	1
sum	46		47

phrase is repeated, namely P DET ADJ N P DET ADJ N. This indicates that the domain of complex disfluencies can be decided explicitly, because the number of words involved is limited.

6.1.2 Complex Disfluencies in NP

In total, seven types were used to cover the occurrences of NP disfluencies. Among them two NP constructions, NP_2 and NP_4, were more often

produced than the other alternative constructions. Interestingly, the most frequent constructions in NP disfluencies are similar to those in PP disfluencies with regard to the remainder structure after prepositions.
NP_2: NP => DET + N (PP_2: PP => P + DET + N)
NP_4: NP => DET + ADJ + N (PP_5: PP => P + DET + ADJ + N)

The distribution of occurrences of NP disfluencies is shown in Table 6.2. 36 out of 45 repaired NP_2 disfluencies and 39 out of 44 NP_4 disfluencies were produced in the sequence of one problem phrase and one corrected phrase. Only a small amount of the repaired NP disfluencies were realized within one single phrase, namely 3 out of 45 repaired NP_2 and 5 out of 44 repaired NP_4 disfluencies. This distribution is in accordance with the results of the overall syntactic analysis in Chapter 3 with respect to the length of disfluencies.

6.1.3 Structural Patterns of Typical PP and NP Disfluencies

This section is concerned with the structural similarity between complex disfluencies produced within PPs and NPs. It is not realistic to expect that a single system of description can cover all the occurrences of speech disfluencies collected in a corpus, because the production of speech disfluencies is not only speaker-dependent, but also situation-dependent. Possible factors influencing the form of production can be language-specific patterns and specific tasks which are assigned to the subjects under an experimental circumstance.

There is a remarkable similarity in the structural patterns, comparing disfluencies produced in PPs with those produced in NPs. Focusing on the most frequent types of occurrences within PP_2 and NP_2, as well as PP_5 and NP_4, one can model this phenomenon as illustrated in Figure 6.1 and 6.2. Apart from the prepositions at the beginning of PP_2 and PP_5, the internal constructions within NP_2 and PP_2 and within NP_4 and PP_5 are identical,

Table 6.2: Occurrence of NP_2 and NP_4 in BAUFIX

DET + N		**DET + ADJ + N**	
repaired		repaired	
(DET, **DET**, N)	36	(DET, **DET**, ADJ, N)	20
(DET, **DET**, *N-d*)	2	(DET, ADJ, **DET**, ADJ, N)	9
(*DET-d*, **DET**, N)	1	(DET, *ADJ-d*, **DET**, ADJ, N)	5
(DET, *N-d*, **DET**, N)	3	(DET, ADJ, *N-d*, **DET**, ADJ, N)	1
		(DET, ADJ, N, **DET**, ADJ, N)	2
		(DET, *N-d*, **DET**, ADJ, N)	1
		(DET, **DET**, ADJ, **DET**, ADJ, N)	1
(DET, *N-d*, N)	3	(DET, *ADJ-d*, *ADJ-d*, ADJ, N)	1
		(DET, ADJ, ADJ, N)	1
		(DET, ADJ, *N-d*, N)	1
		(DET, ADJ, ADJ, *N-d*)	1
		(ADJ, DET, ADJ, N)	1
only interrupted		only interrupted	
(DET, *N-d*)	6	(DET, ADJ, *N-d*)	3
sum	51		47

following the linearly organized surface form. By means of more structural illustrations, as shown in Figure 6.1, a clear pattern repetition can be observed. The loop in the case of NP_2 is placed on the position of DET back to the first element in the phrase which is in this case DET as well. Applying this rule to PP_2, namely the path from DET back to the first element of the phrase, which in this case is the determiner P, we obtain the structure of PP_2. The role DET plays as a position for starting retracing in this rule can be empirically proved. The most frequent position of interruption in the repaired disfluencies is after the determiners.

By focusing on the position of DET and on the initial position of the

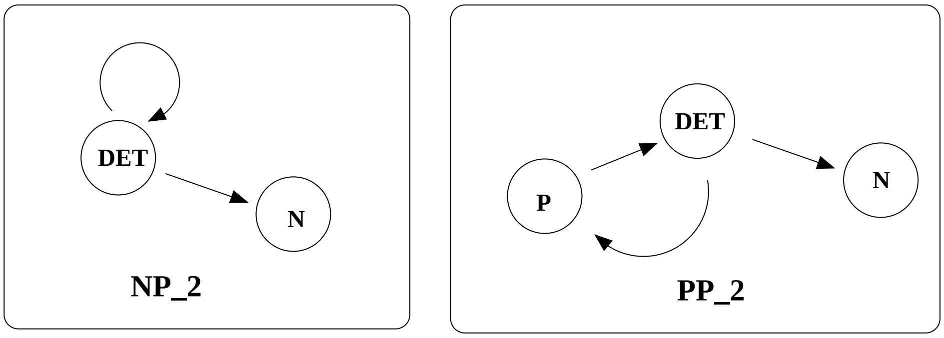

Figure 6.1: Most Frequent Occurrences in NP_2 and PP_2

phrase, as illustrated in Figure 6.2, the same structural similarity between the sequences DET DET ADJ N and P DET P DET ADJ N can be analogously described as in the previous case. The only difference is the added element ADJ between DET and N.

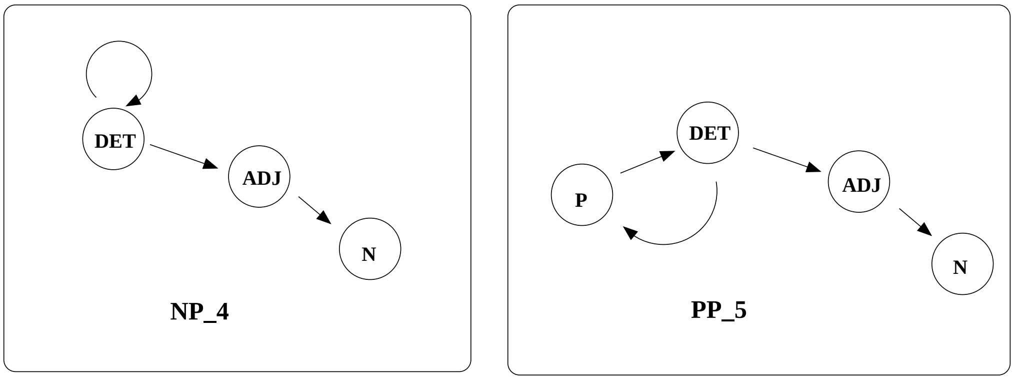

Figure 6.2: Most Frequent Occurrences in NP_4 and PP_5

The result of the syntactic analysis in Chapter 3 is that in addition to the phrasal boundaries we used to identify the disfluencies, only two additional phrasal boundaries were observed within a disfluency in most cases. That is to say, the length of disfluencies does not vary arbitrarily, but rather in a regulated range. More specifically, a repaired disfluency often contains one problem phrase and one corrected phrase. Therefore, the number of words in disfluencies realized in prepositional and noun phrases is rather regular, as well as the part of speech information of the repeated words in the corrected phrases. The part of speech of the initial word in the corrected phrase is often identical with that of the initial word in the previous problem phrase. Interruptions are much more frequently located after prepositions and determiners than elsewhere. To summarize the observations regarding the structural similarities within PPs and NPs, we can first extend the phrasal structure DET N to DET ADJ N and combine these two sequences for NPs into one single pattern, as shown in Figure 6.3. Similarly, for the typical PP disfluencies, a combinatory pattern can be analogously built, as illustrated in Figure 6.4.

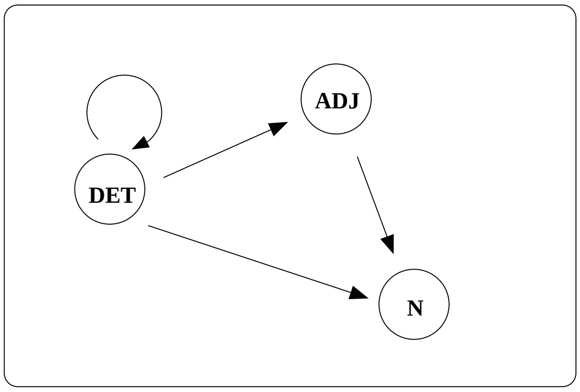

Figure 6.3: Combinatory Pattern Description of NP Disfluencies

6.2 Descriptions of Disfluencies

In Chapter 3, we specified the general segmentation of speech repairs (which are considered in this thesis as a group of complex disfluencies) to a more specific structure, namely from *reparandum, editing terms, repair* to *problem phrase, editing phase, corrected phrase.* Starting with this peculiarity, a structural description of speech disfluencies is attempted in this section.

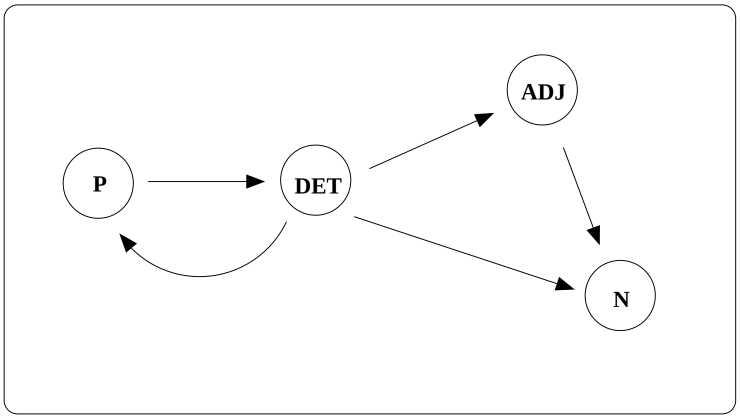

Figure 6.4: Combinatory Pattern Description of PP Disfluencies

6.2.1 Disfluency-Phrase (DFP)

It is not possible to adopt a grammar for written language to describe the structure of utterances in spoken language, because of the intrinsic differences of production and comprehension of written and spoken language. A syntax-based description, nevertheless, is a good alternative to satisfy the need of a systematic formalism and at the same time, to extend the formalism to applied linguistic research fields such as computational linguistics. The unit for this kind of description is called a *phrase*. The reason why the phrasal unit is more adequate for this description is that phrases embody more syntactic potential than other units such as lexical, semantic, pragmatic or prosodic ones. Phrases also provide a convenient way of constructing parsing rules, when the detection and correction of disfluencies are concerned. Another important factor is that the structural description can be more clearly distinguished from the pragmatic realization description, when the syntax of utterances is clearly defined. So the unit *phrase*, making the formal description of disfluencies more syntax-steady, is chosen, to contrast against the pragmatic units which reflect the realization of speech disfluencies.

Disfluency-phrase (DFP) was introduced in Chapter 3 in an operative way in order to make a simple identification of disfluencies possible. We roughly defined *disfluency-phrases* previously as fragmentary or ill-formed phrases. Consider these "ill-formed" phrases as disfluency-phrases in the context of speech disfluencies. Problems appear when dealing with the recovery of ill-formed phrases. That is, simply to expunge phrasal fragments as adopted in most spoken language systems is not the best way to solve the problem of disfluencies such as *die Dreierlis uh Leiste*[3]. By deleting the fragmentary phrase *die Dreierlis* before the editing term *äh* and leaving the rest *Leiste* as a final result, one has the following problem regarding the comprehension of which parts for the constructor in the co-operation are relevant and should be taken. Without exact specifications, the consequence of this simply deleting procedure is: one does not win more information, since there are more than one type of bars and the information **three-holes** is in this case even more important than the word **bar** itself. Furthermore, even grammatically correct phrases can be problematic, if they are in fact *disfluency-phrase* and carry wrong information. Not all grammatically correct phrases serve the purpose of providing the right information. In the example *die gelbe Mutter äh die orangene Mutter*[4], the yellow nut is a correct phrase, but the speaker in fact wants an orange one instead of a yellow one. So if there is no pattern comparison in terms of the internal structure of disfluencies, information carriers can be inappropriately identified and cause confusion.

To extend the coverage of the definition of disfluency-phrase given in Chapter 3, where disfluency-phrases are phrases containing word fragments, disfluency-phrases will be specified in more detail in this section. To be more exact, disfluency-phrases can be of three possible forms: containing unknown words or items, ellipses or complete phrases.

Case 1. When there are unknown items, the nearest phrasal boundary before the unknown item is the beginning boundary of the DFP. This DFP

[3] the three-holes bar

[4] the[feminine, singular, nominative, definite] yellow nut uh the[feminine, singular, nominative, definite] orange nut

then ends with the unknown item. It is for the time being not of central importance whether the disfluency-phrase is corrected later or not.

Example 6.1
also ***in das dri*** *in das zweite und in das dritte auf das zweite und auf das dritte Loch*[5]
DFP: *in das dri*

Case 2. When a phrase is not complete (excluding cases of fragmentary or unknown words) regarding the phrasal rules given for the given language, it is a disfluency-phrase. These kind of sequences, known as ellipses, can be very often observed in spoken language, where for instance a determiner alone can be a phrase, like in German. Because a large number of disfluencies observed in the corpus contain disfluency-phrases in the form of ellipses, the definition of disfluency-phrase should also account for these types of sequences.

Example 6.2
Ich habe einen Würfel ***mit einer*** *mit einem Gewinde*[6]
DFP: *mit einer*

Case 3. Disfluency-phrases can be grammatically correct and complete phrases as well, under certain circumstances. For instance, in the example *die gelbe Mutter äh die orangene Mutter*, the disfluency-phrase *die gelbe Mutter* is a complete and correct phrase. They can be disfluent to the extent that they are prosodically marked, whereby disfluent intonation or peculiar tonal patterns are verbalized. It is however not always the case that prosodic marking can be acoustically measured or audibly perceived. Often, one has a word sequence of co-ordination construction which is in fact a speech repair. The latter construction should replace the former one.

[5]so in the[neuter, singular, accusative, definite] th in the second and in the third on the second and on the third hole

[6]I have one cube with a[feminine, singular, dative, indefinite] with a[neuter, singular, dative, indefinite] bolt

In order to determine whether a phrase is a disfluency-phrase, sometimes additional prosodic features are necessary. In Case 1 and Case 2, the pitch information of the beginning of the disfluency-phrase is similar to that of the beginning of the oncoming corrected phrase, as found in Chapter 4. Temporal contrast distinguishes the repeated words in the disfluency-phrase and in the corrected phrase. Furthermore, contrastive tonal patterns can stress the occurrence of a speech repair where words are explicitly repaired. In Case 3, this kind of prosodic cue is particularly important to tell the difference between a "normal" phrase and a disfluency-phrase, which is grammatically completely correct, but has to be corrected.

6.2.2 Disfluency Structure (DFS)

Starting from the definition given for determining the length of disfluencies, the beginning and the ending phrasal boundaries of disfluencies are fixed. Concerning the internal structure of complex disfluencies, what need to be determined is the region between the problem and the corrected phrase. The beginning of the editing phase in most cases found in the corpus is immediately after the final phrasal boundary of the problem phrase. Similarly, the end of the editing phase immediately precedes the initial phrasal boundary of the corrected phrase. The majority of complex disfluencies have the typical form (problem phrase, editing phase, corrected phrase).

Therefore, a structural description of complex speech disfluencies, called disfluency structure (DFS) from now on, requires at least three components: problem phrase (RP), editing phase (EP) and corrected phrase (CP). An RP is supposed to be a DFP in most cases[7] and often contains false or incomplete information in the course of producing speech disfluencies. A CP is a phrase containing correct information, in relation to the previous incorrect or inappropriate information, although a CP can very well be a DFP, too. The

[7]Certainly, it is also possible that an RP is a perfect "normal" phrase, whereby the phrasal structure is grammatical and there is no prosodic remarkable marking.

corrected phrases are sometimes not completed, since the speaker interrupts his/her attempt at correction during the process of correction. An EP contains editing terms such as lexically marked strings, pauses or interjections which occur after RP and signal that a disfluency is happening. An EP can be absent or present, as found in the corpus. In fact, the minority of complex speech disfluencies are accompanied by editing terms. Thus, this result does not support the presumption made in Hindle's *edit signal hypothesis* [42] from an empirical point of view. Using these three components, complex speech disfluencies can be described in following sequences:
(RP, (EP), CP)
(RP+x, (EP), CP)
(RP, (EP), y+CP)
(RP+x, (EP), y+CP)
where (EP) means EP is optional and x and y are sequences which do not belong to the problem phrase nor to the corrected phrase. Certainly, occurrences such as (RP, (EP), CP) are the majority of complex speech disfluencies found in the corpus. That means, more than 80% of the overall repaired speech disfluencies can be classified into this group. Disfluencies such those found in Example 6.3 can be frequently found in the corpus, whereby a CP is a complete and grammatically correct phrase. Another possible production form of the sequence (RP, (EP), CP), although it does not occur as often as the first case, is that a CP is a disfluency-phrase, i.e. CP is either not completed or contains only single words, as shown in Example 6.4.

Example 6.3
(RP, _, CP)
*also <u>in das dri **in das zweite**</u> und in das dritte auf das zweite und auf das dritte Loch*[8]
RP: *in das dri*
CP: *in das zweite*

[8]so in the[neuter, singular, accusative, definite] thr in the second and in the third on the second and on the third hole

Example 6.4

(RP, EP, CP)

ich habe die Dreierlis uh leiste noch frei[9]

RP: *die Dreierlis*

EP: *uh*

CP: *leiste*

Sequences such as (RP+x, EP, CP), as illustrated in Example 6.5, were seldom found in the corpus. Occurrences such as (RP+x, (EP), y+CP) could not be found at all in the corpus data.

Example 6.5

(RP+x, EP, CP)

und uh drehst die fest mit einmal uhm mit der orangenen Achtkantschraube[10]

RP: *mit*

x: *einmal*

EP: *uhm*

CP: *mit der orangenen Achtkantschraube*

In contrast to (RP+x, (EP), CP), occurrences of the form (RP, (EP), y+CP) are more frequent. This is because after the speaker has interrupted his/her speech, in the case of no sign of EP, the speaker does not initiate the repair directly, but first produces something else, such as adverbs. Therefore, those sequences are identified into the group of (RP, (EP), y+CP), e.g. Example 6.6. Interestingly, the pitch height of the beginning of y is similar to that of the beginning of the RP.

Example 6.6

(RP, _, y+CP)

jetzt hinter den Flügeln, also am ganz am Ende der Stange[11]

RP: *am*

[9] I still have the three-holes-b uh bar

[10] and uh you fix that with once uhm with the orange eight-edges-bolt

[11] now behind the wings, well in the very in the end of the bar

y: ganz
***CP**: am Ende*

6.2.3 Production Patterns of Disfluencies and Examples

Making use of DFS, i.e. focusing on the surface structure of disfluencies, as mentioned in Table 3.5 in Chapter 3, the phenomena of speech disfluencies can be classified in a more structural way. By varying the parameters in DFS, most of the occurrences of speech disfluencies including both complex and simple ones can be described. Disfluencies can therefore be divided into three subgroups according to their DFS: repaired, interrupted and simple disfluencies. Table 6.3 lists the relationship between the structure of disfluencies and their classification in the literature.

Table 6.3: DFSs and their Classifications

Types of DFS	**DFS**	**Classifications**
repaired	(RP+x,_,y+CP) (RP+x, EP, y+CP)	repairs/repetitions
interrupted	(RP+x,_,_) (RP+x, EP,_)	restarts/speech errors
simple	(_,EP,_)	pauses (filled/unfilled)

When CP is available or in other words produced, then it is a repaired disfluency, as shown in Example 6.7, where **CP** is *auf den untersten Würfel*. In repaired disfluencies, RP is repeated or corrected in CP.

Example 6.7
und ähm jetzt müßtest du die schon <u>auf die auf den untersten Würfel</u> sch öh

schrauben können[12]
***DFS**: ([auf, die, _, _], _, [auf, den, untersten, würfel])*

Interrupted disfluencies are disfluencies in which RP is available with or without EP, but CP is not produced. That means, RP is simply interrupted somewhere within PP or NP without being corrected or further specified. An editing phase can be filled by editing terms after the production of RP, but it is not necessary. In Example 6.8, **die unt** is not corrected or specified at all.

Example 6.8
es kommt <u>die unt</u> wenn man jetzt vom Heck her guckt[13]
***DFS**: ([die, unt],[_],[_])*

Simple disfluencies are in fact defined as normal syntactic phrases with interjections or pauses occurring within the phrases or simply pauses or interjections between phrasal boundaries, as shown in Example 6.9 and 6.10. In addition, lexically marked editing phases frequently found in the corpus are **nee** (no), **also** (well), **genau** (right).

Example 6.9
*und da diese **:p:** diese zwei Schienen als Propeller dazwischen*[14]
***DFS**: ([_],[:p:],[_])*

Example 6.10
*so zu den **äh** zu den Rädern*[15]
***DFS**: ([_],[äh],[_])*

One additional remark concerning interjections in spoken language needs to be made. The realization of interjections varies depending on the customs

[12] and ehm now you must be able to sc uh screw that already onto the on the lowest cube
[13] it comes the unt when one now sees from the rear
[14] and over there these these two rails as propeller between that
[15] so to the[plural, dative, definite] uh to the[plural, dative, definite] wheels

of the speakers or perhaps also on the reflexive reaction of the speakers. It is not an easy task to find out for instance why a speaker realizes *uh* instead of *äh*. A detailed study especially concerning interjections in the German language can be found in [27]. It seems that the group of interjections used in spoken language is language-specific. In PP and NP disfluencies examined here, the interjections used throughout the corpus are **uh**, **äh**, **öh**, **uhm**, **mhm** and **öhm**.

6.3 Disfluency Models

We have identified speech disfluencies according to the definition of problem and corrected items in the co-operative dialogues and this criterion seems to be useful in systematizing speech disfluencies. Focusing on the specific type of disfluencies (*problem phrase*, *editing phase*, *corrected phrase*), which are the majority of disfluencies produced in the BAUFIX corpus, the development of exploring "phrase-basedness" in the production of disfluencies, which is at the same time development of disfluency models, is discussed in this section.

6.3.1 Disfluency Relation

This section is mainly concerned with disfluency models, showing different focuses on the structure of complex disfluencies, as illustrated in Figure 6.5. Disfluency models which are to be introduced here deal with internal structural modelling of disfluencies. In disfluency model 1, the relationship between the reparandum and the repair is not clear. As in most of the speech repair research, the focus is only on problem items and corrected items. Sometimes, the initial and final phrasal boundaries at the beginning and end of a disfluency, the interval [**a**, **b**], are taken into account. They determine the length of the disfluency which contains problem and corrected items. The recognition of problem and corrected items is the most decisive criterion in judging whether it is a disfluency or not. Thus, this elementary disfluency model defines a phrase-oriented internal relation within disfluencies.

In Chapter 3, the results of the syntactic analysis determined the internal structure of complex disfluencies. The dots between the problem and corrected words in disfluency model 1 can be specified furthermore by introducing phrasal boundaries between the problem phrase and the corrected phrase. In this step, the internal structure of complex disfluencies shows regular occurrences from the point of view of syntactic phrasing. Where a problem phrase ends does not necessarily correspond to where a corrected phrase begins. There are two additional phrasal boundaries within complex disfluencies which can determine the internal structure of complex disfluencies, namely **c** and **d** in disfluency model 2.

The interval [**a**, **c**] is the problem phrase in which the problem item is located, whereas the interval [**d**, **b**] is the corrected phrase which contains the corrected items. The internal structure of complex disfluencies is more specifically structured in disfluency model 2 than in disfluency model 1. Accounting for phrasal boundaries and syntactic characteristics such as part of speech and linguistic categories of items within disfluencies, the length of a complex speech disfluency is seldom longer than 10 words. The reason for this is that long speech disfluencies are in most cases prepositional phrases and more than two adjectives are seldom used in a prepositional phrase. Furthermore, the end of a corrected phrase is often phrase-final. In other words, ellipses are seldom realized within this type of complex speech disfluency.

The results of the acoustic-prosodic analysis and the syntactic re-analysis contribute to a clearer picture of the internal structure of speech disfluencies, namely that the internal structure can not only be structured by syntactic means, but also by means of prosodic marking. In Figure 6.5, the contents of the intervals [**a**, **c**] and [**d**, **b**] can be more clearly described by focusing on the prosodic relationships between the problem phrase and the corrected phrase. There are in total three prosodic relationships which are used to mark the

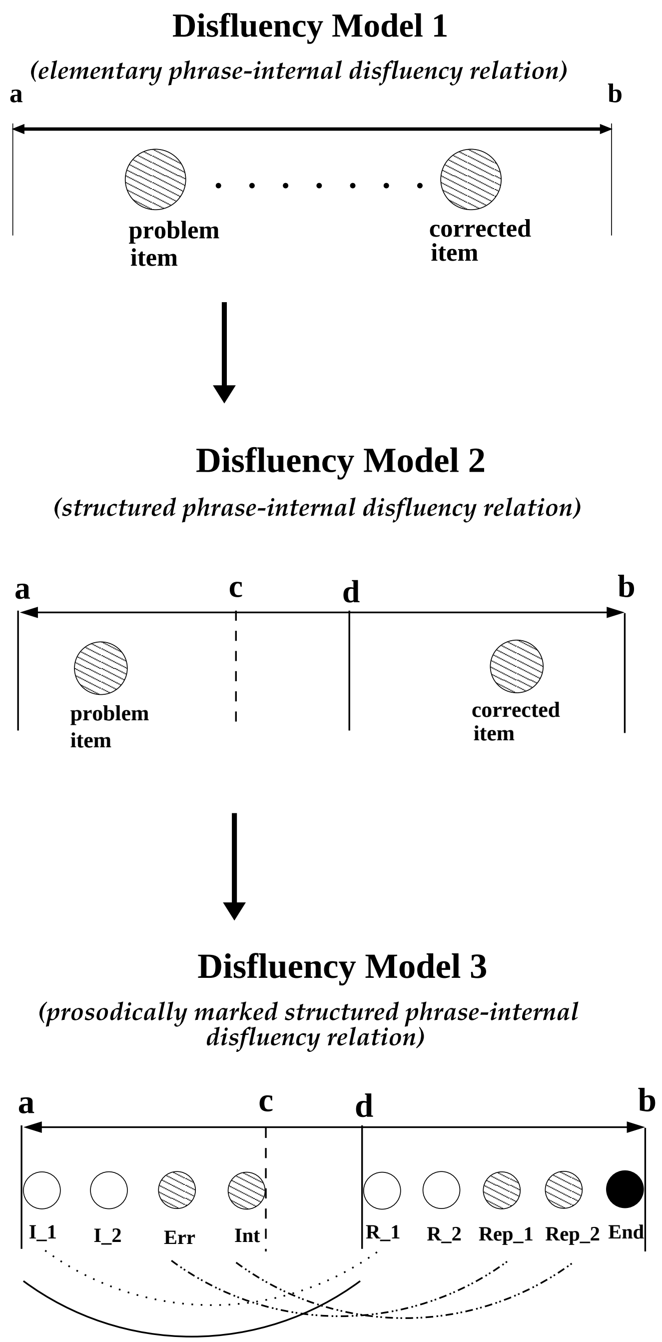

Figure 6.5: Internal Structure of a Complex Disfluency

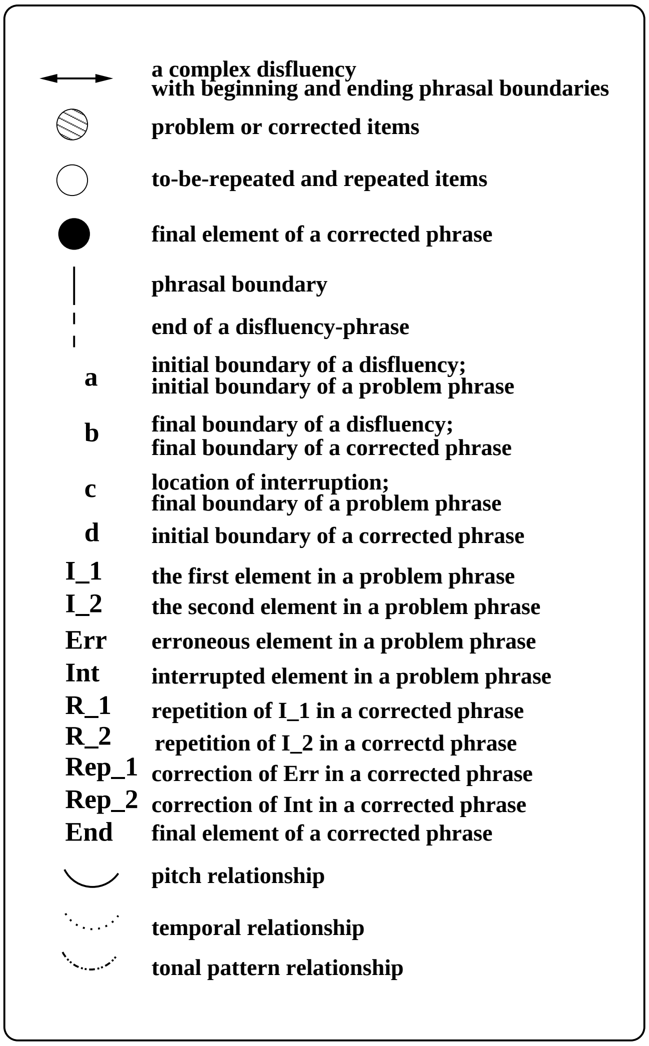

Figure 6.6: Definitions of Symbols in Figure 6.5

internal structural characteristics of complex disfluencies: pitch relationship, temporal relationship and tonal pattern relationship.

I_1 (the first element in the problem phrase) and **R_1** (the first repeated element in the corrected phrase) are usually both phrase-initial and share similar prosodic features. The pitch properties and tonal patterns of **I_1** and **R_1** seem to be similar and show no contrast from the prosodic point of view. This indicates the phenomenon of an intonational restart, when the corrected phrase is started. Regarding the temporal features, **R_1** is significantly shorter than **I_1**. This temporal marking serves the same function as the intonational restart, but from a temporal aspect. Between **I_2** and **R_2**, no particular specification was found in the acoustic analysis, neither in pitch qualities nor in durational properties. Often, contrastive tonal patterns of **Err** and **Rep_1** as well as of **Int** (interruption) and **Rep_2** are used to signal the process of correction, i.e. L-H or H-L contrast. Tonal pattern relationships emphasize the importance and relevance of the ongoing corrected words.

Following the considerations concerned with the three disfluency models, it has been made clear that the internal structural disfluency relations can be specified by syntactic and prosodic means. Furthermore, this regularity can be observed in most of the disfluencies produced, indicating that there exist patterns which speakers prefer or orient to, when they produce disfluencies.

6.3.2 Prosody: Parallel to Syntax

As shown above, acoustic-prosodic features can be found parallel to syntactic ones. In Chapter 4, interruptions often occurred between the problem and the corrected phrases. That is, immediately after the problem phrases are interrupted, new phrases are initiated to repair the interrupted phrases. Thus, the location of interruption is an important site marking the occurrence of speech disfluencies. Not all speech disfluencies identified in the corpus were produced with clearly recognizable acoustic cues, for instance silent/filled pauses or lexically marked sequences. There may possibly exist other kinds of acoustic evidence signalling interruptions such as glottalization [84] which was not investigated in the previous acoustic-prosodic analysis.

The phenomenon of *retracing* outlined in the previous section can be prosodically marked by temporal and intonational means: 1) the first repeated word in a corrected phrase is shorter than the first word in a problem phrase and 2) the onset of a problem phrase is intonationally similar to the onset of a corrected phrase. In Figure 6.7, these two features can be observed between **I_1** and **R_1**. Pitch and temporal relations determine the prosodic relationship between the to-be-repeated and repeated words.

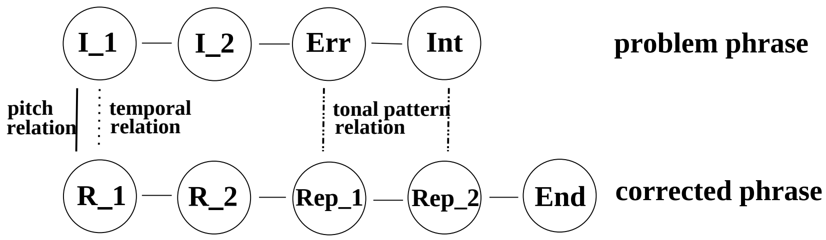

Figure 6.7: Prosodically Marked Disfluency Relations

Similarly, the process of *repeating* and *correcting* can also be prosodically signalled, especially by means of tonal differences: 1) to-be-corrected words in a problem phrase are often realized in a contrastive tone from that of the corrected words in a corrected phrase and 2) to-be-repeated words in a problem phrase are often realized in a similar tone to that of the repeated words in a corrected phrase. As shown in Figure 6.7, the correction of the erroneous or interrupted words can be emphasized by tonal pattern relations between **Err**, **Rep_1** and **Int**, **Rep_2** respectively.

6.4 Finite State Modelling

Significant structural characteristics of repaired disfluencies are to be modelled in terms of finite state devices [61]; [56]; [62]; [3]; [76] in this section. As it will later become clearer, important features in the production of repaired disfluencies can be mapped into mechanisms operating in finite state models.

6.4.1 Finite State Automata

To take the NP construction DET ADJ N as an example, the non-deterministic finite state automaton M_1 with ϵ-transitions is denoted by a quintuple $<$Q, Σ, δ, q_0, F$>$ defined as follows and models the occurrences of NP disfluencies of this construction, where
Q = $\{q_0, q_1, q_2, q_3, q_f\}$,
Σ = {det, adj, n, det-d, adj-d, n-d, ϵ},
q_0 is the start state,
F = $\{q_3\}$ and
$\delta(q_0$, det)=q_1, $\delta(q_1$, adj)=q_2, $\delta(q_2$, n)=q_3, $\delta(q_0$, det-d)=q_f, $\delta(q_1$, adj-d)=q_f, $\delta(q_2$, n-d)=q_f, $\delta(q_f, \epsilon)=q_0$, $\delta(q_1, \epsilon)=q_0$, $\delta(q_2, \epsilon)=q_0$, $\delta(q_3, \epsilon)=q_0$.

The transition function δ can also be formulated in the form of a table as shown in Figure 6.8 which can more effectively show the devices built into the model to simulate the structural characteristics of repaired disfluencies.

Five states are needed, where q_0 is the only one initial state allowed and q_3 is the only state contained in the set of final states F. Six input symbols are used in this model. "Det, adj and n" denote complete words of the categories "determiner, adjective and noun", whereas "det-d, adj-d and n-d" denote incomplete or erroneous words of the categories "determiner, adjective and noun". Det, adj and n reflect the underlying structure of this group of NP disfluencies. In the production of repaired disfluencies, word fragments indicate the occurrence of disfluency and the eventual coming repairs. Det-d, adj-d and n-d therefore represent word fragments which will be delivered back to the start state q_0 through ϵ.

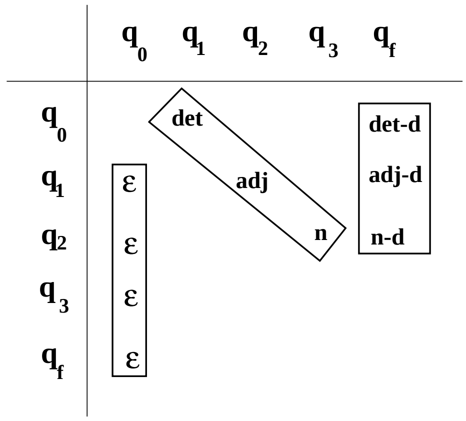

Figure 6.8: Transition Function δ of M_1

The finite state automata M_1 can then be graphically illustrated in Figure 6.9. M_1 accepts for instance the sequence *(det, adj, det, adj, n)* through the path q_0-q_1-q_2-q_0-q_1-q_2-q_3. The sequence *(det, adj-d, det, adj, n)* is also accepted by M_1 through q_0-q_1-q_f-q_0-q_1-q_2-q_3. But a sequence such as *(det, adj-d, adj, n)* which is a repaired disfluency occurring within one phrase is not accepted by M_1.

The pattern of the transition function δ shown in Figure 6.8 can be applied for modelling other groups of repaired disfluencies, by adjusting the number of states and the symbols. For instance, to model the occurrence of PP disfluencies in the construction P DET ADJ N, one more state q_4 and one more symbol p (for preposition) are needed, as well as a transition function which is built similar to that of the finite state automaton M_1 above. The corresponding finite state automaton M_2 is $<Q, \Sigma, \delta, q_0, F>$, where

Q = $\{q_0, q_1, q_2, q_3, q_4, q_f\}$,

Σ = {p, det, adj, n, p-d, det-d, adj-d, n-d, ϵ},

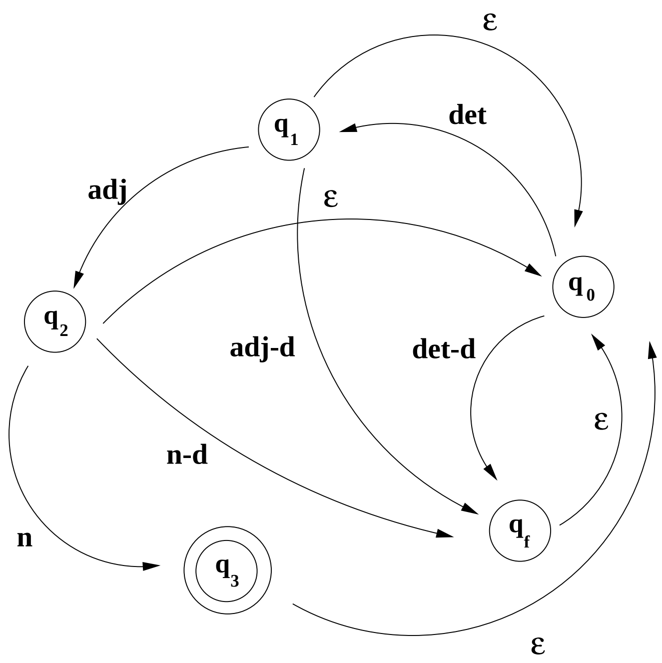

Figure 6.9: Finite State Model M_1

q_0 is the start state,
F = $\{q_4\}$ and
$\delta(q_0,\ \mathrm{p}){=}q_1$, $\delta(q_1,\ \mathrm{det}){=}q_2$, $\delta(q_2,\ \mathrm{adj}){=}q_3$, $\delta(q_3,\ \mathrm{n}){=}q_4$, $\delta(q_0,\ \mathrm{p\text{-}d}){=}q_f$, $\delta(q_1,\ \mathrm{det\text{-}d}){=}q_f$, $\delta(q_2,\ \mathrm{adj\text{-}d}){=}q_f$, $\delta(q_3,\ \mathrm{n\text{-}d}){=}q_f$, $\delta(q_f,\ \epsilon){=}q_0$, $\delta(q_1,\ \epsilon){=}q_0$, $\delta(q_2,\ \epsilon){=}q_0$, $\delta(q_3,\ \epsilon){=}q_0$, $\delta(q_4,\ \epsilon){=}q_0$.

Figure 6.10 illustrates the transition function δ in the form of a table whose pattern is very similar to that of Figure 6.8. Therefore, as illustrated in Figure 6.11, the structure of M_2 is likewise similar to that of model M_1 by applying the same principle.

	q_0	q_1	q_2	q_3	q_4	q_f
q_0		p				p-d
q_1	ε		det			det-d
q_2	ε			adj		adj-d
q_3	ε				n	n-d
q_4	ε					
q_f	ε					

Figure 6.10: Transition Function δ of M_2

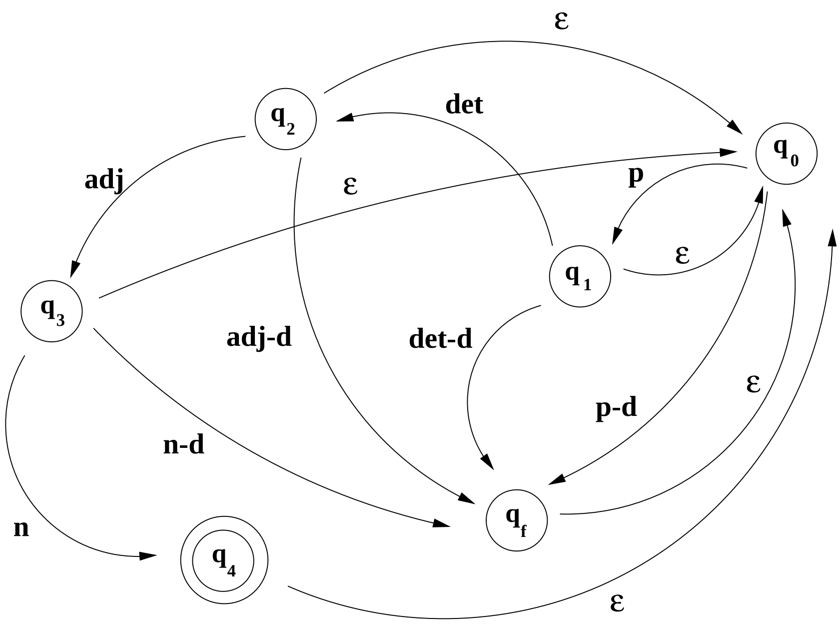

Figure 6.11: Finite State Model M_2

6.4.2 Mapping Disfluency Relations onto FSA

The internal relations within disfluencies mentioned earlier form the internal structure of disfluencies. To sum up the structural relations in disfluencies, there are three important ones, namely 1) the phenomenon of retracing: where the POS of the first word in the corrected phrase is identical with that of the first word in the problem phrase, 2) an interruption is often located after interrupted or unknown words and 3) the corrected phrases are often complete phrases, that is, both prepositional and noun phrases end up with nouns. These relations can be clearly reflected in finite state automata and therefore be formally modelled. Additionally, the internal structural relation also indicates the limitation of the number of words produced in one disfluency. That is to say, immediately repaired disfluencies have a limited number of words and, furthermore, given a certain type of phrasal construction, the number of words is frequently regular. This section is thus concerned with the correspondence between the disfluency relations and their mappings in finite state modelling.

The blocks **det adj n** and **p det adj n** in Figure 6.8 and 6.10 determine the type of phrasal constructions of disfluencies and therefore decide the production possibilities of disfluencies. This fact is significant, if one adopts the pattern-matching approach in recognizing disfluencies, because the underlying structure of phrases, from the syntactic point of view, is decisive in the variations of patterns. Therefore, the number of necessary elements of a transition function is fixed to be $3n + 1$, if the number of elements of the underlying phrasal structure is n, as shown in Figure 6.8 and 6.10. The phenomenon of retracing is modelled in terms of paths going back to the start state q_0, directly or through q_f, by assigning the null path value ϵ, as shown in Figure 6.9 and 6.11. This can be made clearer by looking at the tables in Figure 6.8 and 6.10. The transition functions δ in M_1 and M_2 have a block of ϵ paths respectively. The number of these ϵ paths is dependent on the type of phrasal construction. The ϵ movements heading back to the start state denote the retracing phenomenon in the production of disfluencies.

The possible sites of interruption are modelled through the symbols **det-d**, **adj-d** and **n-d** in the model M_1 and **p-d**, **det-d**, **adj-d** and **n-d** in the model M_2. The state q_f takes care of cases such as interrupted words or unknown items written in **-d** and goes back to the start state q_0 by carrying the value ϵ. The number of possible **-d** words depends on the type of phrasal construction, too. However, the two finite state models M_1 and M_2 can not account for the case of unknown words which are unrecognizable with respect to their syntactic categories. M_1 and M_2 only apply, when the interrupted or unknown words can be classified into their syntactic categories.

A legal input for models M_1 and M_2 can only end in the states q_3 and q_4 respectively. This fact models the phenomenon that a corrected phrase always ends up with the final word of a complete phrase. At the same time, it also means that in both finite state models corrected phrases are necessary. To obtain models which also account for disfluencies containing only interrupted or erroneous problem phrases, one can vary the number of elements in the set **F** of final states.

To sum up this section, the mapping of disfluency relations onto the devices in finite state modelling is illustrated in Figure 6.12. This approach provides a new perspective on the integration of prosody and syntax in respect of disfluencies. The ϵ paths in the model are not especially effective and economical from the viewpoint of the results of the automata generation. But through them an opportunity is given, because they can be used for acoustic or prosodic signals of interruptions or disfluencies.

6.4.3 Models and Empirical Data

After introducing the finite state modelling of repaired disfluencies in PP and NP, this section is concerned with the coverage of the finite state models with respect to repaired disfluencies produced in the BAUFIX corpus. The goal

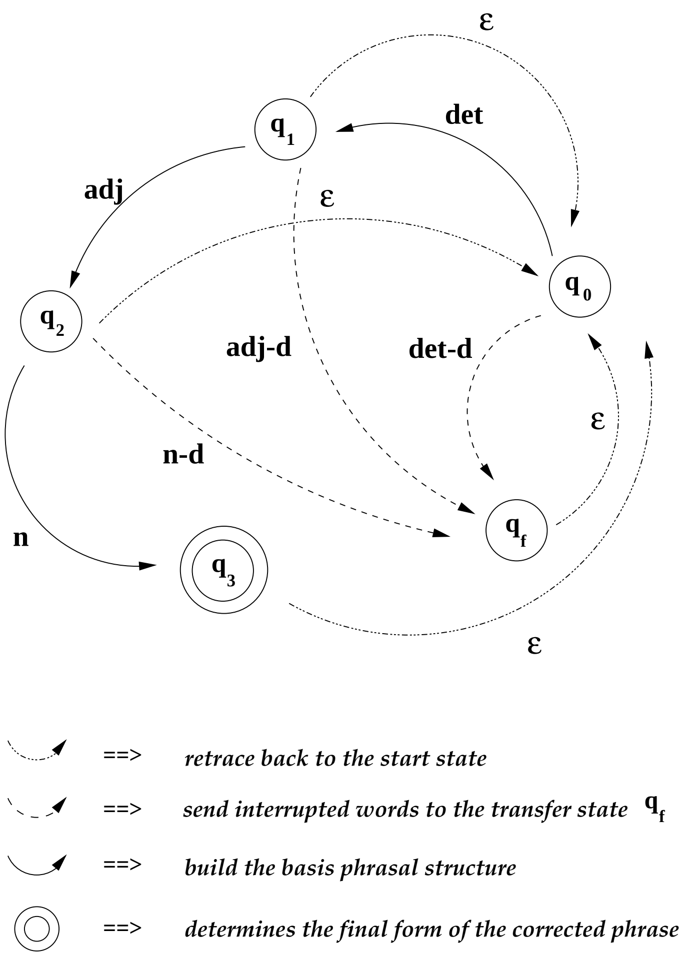

Figure 6.12: Mapping Disfluency Relation onto FSA

of this evaluation is to see if this kind of modelling can model the majority of disfluencies produced in reality and furthermore determine if the production of the repaired disfluencies is regular.

By applying the four principles mentioned above, namely the retracing mechanism, the treatment of interrupted words, the basic phrase structure as well as the single final state, finite state models for other groups of disfluency occurrences can be constructed in a similar way. Models for PP and NP repaired disfluencies can be systematically extended, to include more simple or complicated phrasal structures. For instance, M_2 is in fact an extension of M_1 by simply increasing the number of phrasal elements from three to four. Similarly, by reducing the number of elements of a given underlying phrasal construction to two in the case of noun phrases, e.g. DET N, one gets a finite state model M_3 as shown in Figure 6.13.

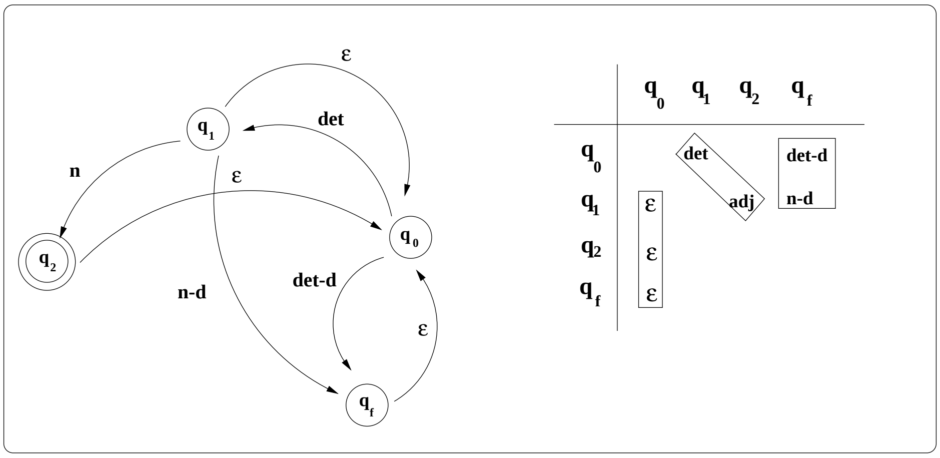

Figure 6.13: Finite State Model for DET N Disfluencies

By repeating similar processes of extension and reduction for all PP and NP construction types listed in Figure 3.4 and 3.6 in Chapter 3, each construction type can be modelled by a finite state model respectively, except NP_6[16]. 78.3% of PP and 83.5% of NP disfluency occurrences found in the

[16]Again, NP_6 contains the **so ein** constructions whose internal structures vary to a

BAUFIX corpus can be covered by this kind of finite state modelling. The evaluation of each construction type is listed in Table 6.4. In the middle column, the number of occurrences of each construction type which can be modelled in finite state models are given. The right column contains the number of occurrences of repaired disfluencies which were found in the BAUFIX corpus.

Table 6.4: Frequency of Occurrences Modelled by FSA

type	# of occurrences (modelled)	# of occurrences (produced)
PP_1:	4	5
PP_2:	35	43
PP_3:	4	4
PP_4:	5	9
PP_5:	35	45
sum	83	106
NP_1:	5	5
NP_2:	39	45
NP_3:	13	16
NP_4:	38	44
NP_5:	6	11
sum	101	121

One important reason why there are occurrences deviating from the "typical" production is that the repairing process is dependent on the form and syntactic position of speech errors made. Some of the complex disfluencies contain interrupted lexical items which are often adjectives and nouns. In this case, the errors were corrected immediately without retracing back to the beginning of the problem phrase. This phenomenon can be further modelled

great extent. In this thesis, this group is not further classified.

by developing a more sophisticated FSA based on that suggested above, by assigning the ϵ paths more specific values to distinguish the difference between types of speech errors or interruptions.

6.4.4 Discussion

The preliminary modelling proposed above focuses more on the internal structure of problem and corrected phrases than on the location and the form of interruption in disfluencies. Only the case of interrupted words has been considered in the models. The models do not account for interruptions such as silent pauses and filled pauses which also frequently occur in spoken language. Acoustic indicators of interruption can possibly be glottalization. For a complete model of disfluencies, those phenomena should be taken into account as well. Furthermore, the probability of frequency of retracing and interruption differ from word category to word category, as well as from one phrase structure to the other. As shown in Chapter 3, it is more possible in certain locations to interrupt or to retrace back to the beginning of a problem phrase. This characteristic can be built into a stochastic model with probability features assigned to the paths in the model. To extend the models suggested in this chapter by adding the empirical results obtained in the syntactic analysis, one can construct a model which is in accordance with the reality of the production of speech disfluencies in a more concrete way.

Another extension possibility of the models proposed is to compose them. It is theoretically possible to combine a number of small finite state automata, say for all construction types of prepositional phrases, and to construct one large finite state automaton to cover all the occurrences in one model. In doing so, however, the significant features in the individual groups of disfluencies with respect to the basic phrasal structure will no longer be clearly presented and differentiated.

6.5 Summary

In Chapter 6, a syntax-oriented description convention was suggested, based on the occurrences of disfluencies produced in the corpus, aiming to cover as completely as possible the variations of disfluency in these occurrences. Using phrasal boundaries as eventual disfluency boundaries has proved to be effective in systematizing and grouping the structure of disfluencies. Determiners, play an especially important role in the principles of pattern description of repaired disfluencies in NP and PP. By fixing the position of determiners as the location of interruption and re-initiating the entire phrase from the beginning of the phrase, one can effectively generate the most complex disfluency occurrences. The reason for choosing determiners as a relevant site for retracing, is that in the corpus analysis the location of interruption often occurs after determiners and this approach has also proved to work well in systematizing NP and PP disfluencies.

Furthermore, notations have been suggested for the structural description of disfluencies including simple and complex ones, independent of individual phenomena of disfluencies such as speech repetitions or speech repairs. The term disfluency-phrase (DFP) used in Chapter 3 has been further specified and a new notation disfluency structure (DFS) accounts for structural variations in the production of disfluencies found in the corpus. Disfluency models suggested integrate syntactic and prosodic information and internal structural relations within complex disfluencies. Important structural relations within complex disfluencies such as the retracing of the corrected phrase, the interrupting of the problem phrase and the completing of the corrected phrase are then modelled in finite state models. Non-deterministic finite state automata with ϵ movements clearly formulate these three features in terms of finite state devices. The models and the empirical data found previously were compared. Furthermore, these models were seen to cover the majority of the occurrences found in the corpus.

Chapter 7

Characterizations of Spoken Dialogues

The role syntax plays in written language is not necessarily equivalent to that in spoken language. Syntactic knowledge and rules, nevertheless, influence the production of spoken language to a great extent, as previously shown in the case of speech disfluencies. It is unlikely to examine the grammar of spoken language without taking into consideration prosody of speech and the interaction between interlocutors. Both interactional and prosodic influences have essentially to do with temporal features of speech production, since the organization of utterances in spoken language is no longer only dependent on the producer of the discourse exclusively, but also on the addressee and his reaction. This dynamic interaction between interlocutors causes unwanted turn-taking, interruption and resumption. To investigate these phenomena, speech disfluencies provide an advantageous opportunity, since one obtains access to the usage of spoken language through focusing on how various linguistic means are adopted under specific circumstances.

This chapter is concerned with selected significant characterizations of interactive spoken dialogues, from the point of view of the production of speech disfluencies. Specifically, the relationship between syntactic and prosodic units and interactional factors will be discussed. Complex disfluencies such as speech repairs can clearly illustrate this correlation from the structural perspective. The focus will be on interruptions and resumptions occurring within complex disfluencies in spoken dialogues. Complex disfluencies, furthermore, show the simultaneous usage of syntactic and prosodic means to emphasize interruptions and resumptions which are often produced in interactional spoken dialogues. In addition to the conventional linguistic means to segment discourse structure such as syntactic and prosodic means, disfluency markers are grouped systematically with regard to their contribution in organizing the local structure of spoken discourse and will be discussed in detail in the last section of this chapter.

7.1 Grammar and Spoken Dialogues

As shown in the empirical studies earlier in this thesis, ungrammaticality in spoken discourse does not mean that grammar no longer plays an important role in the production of spoken language. On the contrary, ungrammatical sequences indicate particular grammatical positions or rules which the speaker prefers to begin or interrupt. Take speech repairs as an example. When speech repairs are produced, their realization is strongly syntax-oriented. The production of an immediate speech repair follows and adopts grammatical segmentation, namely phrasal segmentation. This section is therefore concerned with the relationship between syntactic characteristics and the production of spoken dialogues, focusing on the informational structure of spoken dialogues with respect to their syntactic distribution.

7.1.1 Interaction or Grammar?

With regard to syntax in spoken dialogues, syntactic rules contribute to the discourse organization to the extent that the syntactic knowledge of a speaker about his language influences his choices, especially reflexive choices, of the language form he uses to react under certain circumstances. Within a very local and small domain, one can predict specific structure which are most likely to be used for certain communication situations, for instance speech repairs. When speech errors are to be repaired and an interruption is available, then it is possible to predict what and how will come next, rather independently of interactional factors. The construction of the surface form of speech is for the most part predictable, as well as the location of an interruption. It is therefore not completely true to say that syntax is merely responsible for written language and interaction is most important in spoken dialogues.

However, regarding the surface structure and the information structure of a spoken discourse in conversation analyses, interactional factors are very often the essential criteria for segmenting spoken discourse. Turn-taking, moves and acts focus more or less on changes of speakers, references and topics, respectively. These changes are often identified in dialogues without starting from the consideration of the surface structure, that is, without directly taking into account the related syntactic information structure. Under both semantic and pragmatic perspectives, action and reaction of dialogue participants mainly create the guiding structure of the discourse [96]; [103]. For instance, Gülich & Kotschi [37] mentioned reformulating activities in oral communication, focusing on the aspect of discourse production. From the point of view of formulating procedures, two reformulating results are distinguished: paraphrastic and non-paraphrastic reformulating. They principally correspond to the phenomena of repetitions and corrections considered in earlier chapters. Three aspects of corrections were analyzed by Gülich & Kotschi: corrections of form, corrections of formulation and correction of contents. Syntactically oriented structure was not explicitly applied as criterion to classify groups of corrections, as they primarily focused on the formulating procedures of the speakers. The syntactic aspect of speech production was not the most decisive factor in their approach.

There is no doubt that an intersection of functions of interactional and syntactic factors is present in the production of spoken discourse. Schegloff [101] and Ford & Thompson [28] discussed this intersection of interaction and grammar from the aspect of turn organization and interactional units in conversations respectively. The interaction among dialogue participants directly determines the form and result of the dialogue. With respect to the linguistic competence of speakers, one can not ignore the influence of syntax in composing utterances, i.e. the main components of dialogues. To take speech disfluencies as an example, speech disfluencies can be classified according to their interactional functions, such as hesitating, interrupting or repairing. These functions at the same time reflect the process and sequence of dialogues. On the other hand, however, features found at the surface structure make the grouping and structuring clearer.

7.1.2 Syntactic and Prosodic Units

It is proposed in this thesis that structural descriptions, especially for spoken language should be made available. These descriptions do not necessarily to be identical with or totally different from the ones for written language. This attempt, especially for German, has been realized by Fiehler [26], Kindt [52] and Rath [94]. Spoken and written language are principally based on a similar syntactic foundation, whereas signals sent by speakers and listeners (Sprechersignale and Hörersignale) influencing the structures of utterances, show that utterance units do not always correspond to sentences. Depending on the type of spoken discourse, the deviation between utterance units and sentences can be enormous [94]. The corresponding prosodic units also vary to the extent that utterance prosody such as intonation and melody no longer makes sense, if from the syntactic point of view, utterances are segmented into units which are temporally too long or whose words are too many.

A definition of sentences in spoken language was suggested by Kindt [52] which is based on the correctness of spoken sentences. He also used ellipses and repairs to argue that they are in fact sentence-internal phenomena which should not build independent units. According to the results obtained in the empirical studies, speech repairs deserve a detailed description which is not sentence-independent, but sentence-internal and, as previously shown, phrase-dependent. The so-called syntactic repairs were also prosodically marked. Whether the to-be-repaired errors were lexical or syntactic, prosodic markedness in terms of temporal and tonal markings were found. This fact supports the notion that prosody plays a different role than other linguistic areas such as phonological, lexical or syntactical ones. Although syntactic repairs can be more or less easily recognized at the surface level, completely ignoring the prosodic marking can also be problematic. Take the complex disfluency "*dieses Rad kommt also direkt nach der gelben Mutter aeh nach der gelben Schraube*"[1] as an example. Without *aeh* and the intonational reset at the beginning of the phrase *nach der gelben Schraube*, it would be difficult to tell whether it was a repair or a conjunction construction.

7.2 Describing Spoken Dialogues

Syntactic, prosodic and interactional factors all influence the structure of spoken language to a certain extent. Syntactic features are mainly responsible for the organization of phrases and utterances, whereas prosodic means are more often used to emphasize and express additional intentions and extraordinary syntactical constructions. Interaction takes place all the time in spoken dialogues, influencing the reaction of interlocutors and their linguistic realizations. These three factors are to be discussed in this section.

[1] this wheel is located directly after the yellow nut aeh the yellow bolt (Sagerer *et al.* 1994)

7.2.1 Syntactic Phrasing

There is no need to stress the importance of syntax in composing sentences or utterances in written language production. However, the importance of syntactic factors in spoken language is often neglected. *Phrases* possess particular functions in composing utterances in spoken language. On the one hand, phrases are constituents of utterances. On the other hand they consist of lexical words which again are the basis for determining utterance meaning. By investigating complex disfluencies, the relationship between syntax and prosody was established to the extent that syntactic phrasing is conforming with prosodic phrasing. Based on this result, this section proposes a central and organizing role of phrases in spoken language, in particular, in spoken German. Further clarifications of the role of phrases in spoken language are presented below.

Phrases as units Phrases have proved to be useful units in describing the phenomenon of complex disfluencies, to the extent that systematic results were obtained by adopting phrasal boundaries as disfluency boundaries. In other words, the production of complex disfluencies, speech repairs included, is strongly phrase-based. Phrasal boundaries not only mark boundaries of ordinary phrases and appositive phrases, but boundaries of complex disfluencies as well. Determining utterance boundaries in spoken discourse is apparently a difficult task, because it is not always the case that punctuation used in written texts can be re-found in spoken discourse in the form of prosodic pauses and intonation. Therefore, deciding on the beginning and the end of utterances in spoken dialogues is not very effective, for the reason that re-initiations of utterances, so-called restarts, take place very often, depending on the reactions of the listeners, or on the ongoing attempts of the speakers. Especially in task-oriented dialogues, where additional appositive phrases are used to clarify the descriptions of objects, the distinction between initiated appositive phrases, re-initiated repairs and restarted utterances is often not clear. To adopt the term phrase as a unit instead of "sentence" or "utterance" from the syntax-oriented point of view or "move" or "turn" from the pragmatics oriented point of view, provides better access in dealing with these sequences.

Context of phrases In contrast to the explicit rules of internal structure of utterances, the contextual information of phrases is much more essential. That is, to develop rules for possible contextual phrases can be more useful than to give rules of utterances, in the sense that the utterance boundaries should be explicitly determined. Take the utterance "*also Du hast oben die Tragflächen, in der Mitte die Fünferstange, dadrunter kommt die die Dreierstange zwa und zwar wird die genauso ange oeh befestigt wie hinten, also dass das eine nach vorne steht dann*"[2] as an example. It is extremely difficult to clearly set up the boundaries between utterances, even with the help of prosody.

7.2.2 Prosodic Phrasing

Sanderman [98] has discussed the relationship between perceptual boundary strength (PBS) and pauses, melodic boundary markers, declination reset and pre-boundary lengthening by running perception experiments. From the point of view of speech production, this thesis explored a similar issue as in [98]. Prosodic phrasing can be realized by producing pauses, by F_0 reset or by melodic boundary tones. Prosodic information is no doubt attributable to structuring discourse and is used by listeners in segmenting utterance units. As Goldman [34] pointed out, the temporal characteristics of pauses are related to syntactic structures. The relationship between prosody and spoken discourse is similar to that between syntax and written texts. Brazil [7] and

[2]so, you have the wings on the top, the five-bar in the middle, below it comes the the three-holes-bar zwa and then it will be exactly ange oeh fixed as behind, so that that is located forwards then (Sagerer *et al.* 1994)

Coulthard [15] explored the significance and function of prosody in discourse meaning, whereas Hirschberg & Grosz [43] and Hirschberg & Pierrehumbert [44] concentrated more on the relationship between prosody and discourse structure. To consider the relationship between prosody and spoken discourse from another point of view, namely by looking at the prosodic segmentation of disfluecies in spoken dialogues, this section will concern itself with some prosodic means used to locally structure spoken discourse.

Temporal features That syntactic and prosodic boundaries do not necessarily fall together in spoken language is not a new issue. However, they still have something in common with regard to the function of segmenting utterances. Terken & Collier [121] proposed, supported by their empirical data, that prosodic features of the constituent boundaries between NP and VP are distinguished from those elsewhere. The pitch contours provide separation cues between NP and VP. Furthermore, pauses in this position are often present. Between NP-NP and NP-PP, it is not always the case that the prosodic phenomena above are available like in the case of NP-VP. This shows that prosodic cues are indicative of phrasal combinations with different syntactic properties. Temporal features are also used by Hirschberg & Pierrehumbert [44] to indicate prosodic phrasing such as pauses and the lengthening of the phrase-final syllable. The pre-boundary lengthening too has been suggested with respect to prosodic phrasing within lexical constituents in [125]. According to the results obtained in this thesis, temporal markings sometimes support the function of syntax, focusing on the internal prosodic relations within complex disfluencies. The repeated words in the corrected phrases are temporally marked; they are also significantly shorter than their counterparts in the problem phrase.

Pauses Sometimes long pauses do not imply that there are major constituent boundaries. When a speaker hesitates, there could be a long pause within a phrase and then the phrase is completed after the pause. This kind of long pause does not always signal an utterance boundary. For instance in the data investigated in this thesis, some hesitation pauses are much longer than pauses between clauses.

Melodic features In the investigation done by Swerts & Geluykens [116], the majority of final clauses in the instructions in their corpus data were realized in low-ending contours. This observation supports the notion that syntactic and melodic features have phrasing characteristics in common. Hirschberg & Pierrehumbert [44] indicated occurrences of extra melodic elements in the end of phrases such as pitch range. The pitch range at the end of declarative utterances is lowered. Final lowering seems to be indicative of spoken utterance with respect to speech melody[3]. The phenomenon of F_0 reset in the case of complex disfluencies also supports the notion that prosody is adopted under specific circumstances. When the repair is to be initiated, the melodic features reflect the re-initiation process of the previous problem strings.

Cues phrases In structuring spoken discourse, one has a variety of means to apply. Syntactic structure is the basic structure for understanding. Terms realized in the editing phase between problem and corrected phrases can be used as cue phrases. Cue phrases e.g. *uhm, well* not only indicate hesitation, but pragmatically structure the spoken discourse.

7.2.3 Interaction

In spoken dialogues, interaction under certain circumstances plays a more important role than syntax. Following the dynamic interactional realizations of

[3]Not all results support the final lowering, a counterexample is for instance: Cooper & Sorensen [14] mentions that phrasal boundaries are usually accompanied by a fall-rise pattern in F_0 and the first stressed syllable of the post-boundary constituent usually carries a F_0-peak.

interlocutors is often more effective in understanding speakers and integrating into the speaking community, than merely concentrating on the syntactic constructions on the surface level. Certainly it would be an exaggeration to claim that context and interaction are central and that the syntactic form of language is peripheral. However, interaction does influence the production of speech and the organization of discourses. Interaction, in particular, is clearly observable in the task-oriented dialogues, as used in our corpus, and can give indications of the type of dialogue such as instructional or confirmational further inquiries or disagreement. One can not clearly divide interaction types from their syntactic forms. Regarding syntactic constructions, most of the dialogue acts usually correspond to specific sentence types.

Turn-taking Turn-taking signifies a change in speakers and is, therefore, an important feature for segmenting the information structure of spoken discourses, especially dialogues. In our task-oriented dialogues, turns are in some cases not completely suitable for discourse segmentation because of a variety of overlapping speech stretches and back-channel responses. Very often, the constructor and the instructor spoke simultaneously. That is, one subject of a construction pair often immediately reacted to the statements or questions of the other, while the other carried on his/her own speech. In these cases, it is difficult to decide the beginning and the end of turns. In most times, the instructors said much more than the constructors. Usually, the constructors only asked questions or confirmed the information they received. The role of interlocutors assigned to the constructor and the instructor is clearly defined and can also be clearly distinguished. Interaction units, such as the construction units particularly in our corpus, may be more useful than turns.

Instruction units Due to the specific task assignments in the spoken dialogues investigated previously, the discourse type of these dialogues is different from that of spontaneous or read speech. Thus, markers separating instruction units can contribute to dialogue structuring as well. Conjunctions such as *also*, *dann* or *und* are particularly important.

7.3 Disfluency Markers in Spoken Dialogues

Syntax, prosody and interaction account for the structuring of spoken dialogues. It is difficult and also very time-consuming to directly investigate the relationship between these three factors in spoken dialogues. Speech disfluencies however provide a good opportunity to obtain indications about the production of spoken language. They occur, on the one hand, very often in spoken dialogues; on the other hand, they mark the malfunction of language processing and under such circumstances they expose the applicability by other linguistic or extra-linguistic means. Because of the frequent occurrences of disfluencies in spoken language, if one can identify where the disfluencies occur, it automatically makes the task of structuring a spoken discourse easier, both locally and globally. Similar phenomena such as *discourse markers* have recently been used for the purpose of segmenting and marking spoken dialogues [10]; [38].

In spoken language one can not literally declare problems of message delivery or language processing, whenever something goes wrong. For instance, in the case of local formulation problems, the speaker will seldom choose the alternative, "*Sorry, I have made a mistake. What I wanted to say is in fact* ...", but instead use prosodic and syntactic means to illustrate the problem and correct the problem immediately. More specifically, the speaker begins his/her repairs with a new phrase in a similar pitch height and marks the corrected words tonally. In the lexical formulation, the delivery of speech would be too slow and therefore ineffective. Non-lexical means are more frequently expressed and can be more efficiently used in spoken communication than lexical ones. Prosodic and syntactic forms are the two most frequently adopted means signalling that something is wrong with the already produced speech strings or the to-be-produced strings.

7.3.1 Disfluency Markers

What kind of signals can provide indications in the realization of disfluent speech? Interruptions are the most conspicuous cues in disfluent speech, because while the speech is interrupted, clear acoustic and syntactic cues can be obtained. Lexically marked discourse particles, interjections and silent pauses have the same function as well. Once they are recognized, it is highly possible that there is a speech disfluency. When the speech is immediately corrected, the position of resumption to a great extent contributes to signalling the ongoing process of speech repairs. Location of resumption can be the onset of corrections or repetitions, whereas corrections and repetitions can be distinguished or made clear by prosodic means. In the case that the contents of the editing phase are available, the task of identifying the location of resumption is therefore made easier. These locations are illustrated in Figure 7.1. Furthermore, not only the individual locations of interruption, resumption and the editing phase are directly related to the organization of complex disfluencies, but specific patterns of speech disfluencies are also useful for the identification of disfluencies. The entire phrasal pattern can be regarded as a sort of disfluency marker as well. As shown in Chapter 6, there are specific patterns with regard to the POS sequence, which are more preferred by speakers than other types of construction.

7.3.2 Interruption, Resumption and Correction

Especially in task-oriented spoken dialogues, interruption, resumption and correction are three elements frequently realized. This is due to the fact that the tasks must be carried out precisely and that the subjects have not prepared their instructions previously. Thus errors, interruptions and corrections can and do take place. When these three elements are produced, the

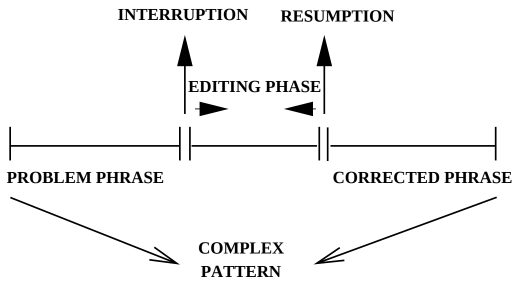

Figure 7.1: Significant Positions for Disfluency Markers

sequence is usually in the order of: interruption, resumption and correction. How can they be marked by linguistic means? What kinds of possibilities do the speakers have to mark these locations? In Table 7.1, two non-lexical means of marking disfluency, prosodic and syntactic ones, are introduced. Disfluency markers related to the interruption can be applied in all kinds of disfluencies containing interruptions, including simple and complex ones. In simple disfluencies, errors are not corrected after the interruption. The other groups of disfluency markers are especially aimed at covering complex disfluencies which include resumption and/or correction.

From the prosodic aspect, an editing phase can be recognized by measurable silent pauses or prolonged vowels in the case of interjections. Filled pauses usually have flat intonational contour. A possible syntactic cue of an editing phase is when unknown items occur; they can be filled pauses or interrupted words. Therefore, unknown items can also refer to word or phrasal fragments which indicates the occurrence of an interruption. Interrupted intonation contour is a signal of interruptions. If the declination of an utterance does not follow its predicted course, it is possible that it is an intonational reset which indicates the location of a resumption. Durational marking of items in the resumed strings can contrast the temporal difference of the previous item with that of the resumed one. Accordingly, the POS of these two items also shows that it is a re-initiation of the previous strings. The process of correction can be effectively marked through tone contrasts.

Table 7.1: Disfluency Markers

	prosodic aspect	**syntactic aspect**
EDITING PHASE	- pauses - prolonged vowels - flat F_0 contour	- unknown items
INTERRUPTION	- interrupted contour	- word fragments - phrasal fragments
RESUMPTION	- intonational reset - temporal marking	- onset of phrases with the same POS
CORRECTION	- tonal marking	
PATTERN		- similar phrasal structure

One will perhaps predict that a similar F_0 contour can also be an indicator of repaired strings. This is only true to a partial extent. The beginning of the corrected phrase shows a similar course to that of the previous problem phrase, but the words which are to be corrected are usually produced in a contrastive tonal contour from that of the previous problem string. So a similar F_0 contour is not a reliable cue to recognize a pattern of complex disfluencies. In contrast, the syntactic aspect is helpful in this case. Speakers tend to adopt the same construction they have used to start their repairs. Therefore, similar phrasal structures can also be a useful disfluency markers.

7.4 Summary

Chapter 7 is essentially concerned with characteristics of interactive spoken dialogues, mainly focusing on the role of syntax, prosody, interaction and disfluency markers. It is obvious that one needs a specific grammar for spoken language. Chapter 7 proposed that not only syntactic phrasing can contribute to the organization of spoken discourse, but prosodic phrasing and interaction factors. According to the results obtained in relation to the production of disfluencies, prosodic features distinguished and disambiguated syntactically difficult and ambiguous constructions, for instance, disfluencies from co-ordination constructions. Disfluency markers signal the occurrence of a disfluency and possible segmentation problems in constructing discourse syntax. Making use of information about where and how interruptions present themselves are is for instance an efficient method in detecting disfluencies. Similarly, prosodic marking signalling the incidences of resumption and correction also serve as disfluency markers.

Chapter 8

Conclusion

Two questions were asked in this thesis:

1) How are speech disfluencies produced from the linguistic point of view, focusing especially on the syntactic and prosodic features?
2) Do regular internal relations exist within speech disfluencies and if so, what do they look like?

Method

These two questions were examined by means of empirical studies and formal models. With respect to the first question, a corpus analysis was carried out to investigate the production of disfluencies from the syntactic point of view. The acoustic-prosodic features of selected types of disfluencies were also examined. A further syntactic analysis was undertaken to secure the syntactic features found in the speaker-independent corpus analysis was refound in the speaker-dependent data which were used in the signal analysis. Furthermore, the empirical results of the corpus analysis and signal analysis should provide empirical evidence for the existence of regular internal relations within speech disfluencies as asked in the second question. Finite state techniques were used for describing this regularity.

Results

The results of the corpus analysis show a highly significant syntactic regularity when disfluencies are produced. Significant results were found with respect to the following features:
1) the linguistic length of disfluencies,
2) the syntactic category of disfluencies,
3) the construction types of disfluencies,
4) the location of interruption,
5) the repair onset and
6) the repair offset.

The results of the signal analysis show that the production of disfluencies is also prosodically marked. Three prosodic means were found to be used for this purpose:
1) the intonational reset of the repair onset to the error onset,
2) the temporal marking of the first word in the corrected phrase and
3) the tonal marking of the corrected words in the corrected phrase.

Furthermore, according to the results obtained by the speaker-independent and -dependent syntactic analyses as well as the speaker-dependent acoustic-prosodic analysis, a clear mapping from the syntactic level to the prosodic level was found. This means that one can find both syntactic and prosodic cues at specific positions within or around speech disfluencies. Syntactic and prosodic effects run parallel to each other.

Thus, the hypothesis that there exist specific phrase-internal disfluency relations has been empirically supported. Based on the empirical results, a formal description system was developed to cover the majority of the disfluencies produced in the corpus. The disfluency relations found were subsequently modelled by means of three disfluency models which also demonstrate the progress of investigating internal structures of disfluencies. Both syntactic and prosodic features were integrated into the final model and the relationship between syntax and prosody with respect to disfluencies was established. These relations were formally described in terms of finite state automata. Only syntactic features were explicitly considered. However, these automata provide an opportunity to integrate prosodic features into a disfluency model.

To summarize the results, this thesis has empirically and formally specified the typical form of disfluencies which are syntactically and prosodically marked. Phrasal boundaries play a decisive role in the realization and completion of speech repairs and prosodic cues have also been found parallel to phrasal boundaries.

Further work

The sequence of the typical complex disfluencies is structurally similar to the mental model development suggested by Johnson-Laird & Byrne [49], namely model construction, model inspection and model variation. What can be asked here is the question, "What is the corresponding syntactic or rather linguistic unit for this cognitive model when language is verbally produced?" Can it be the unit *phrase*?

As we saw earlier in this thesis, prosodic markings were found parallel to syntactic constructions in a very consistent way signalling and emphasizing that there exists occurrences of disfluency. Is prosody therefore useful in recognizing speech repairs and therefore helpful in acquiring correct information? Is redundancy, in the case of prosody in the production of disfluencies, necessary, useful or superfluous?

In computational linguistics, it has been difficult to parse spoken language which contains a variety of disfluencies. By using the syntactic and prosodic features identified in this thesis, as shown in the finite state models, one has another possibility of developing mechanisms for detecting and correcting disfluencies. More specifically, the syntactic and prosodic markings of disfluencies can help identify phrasal boundaries.

Appendix A

Glossary

corrected items: Corrected items are identifiable words or strings of words which are produced to correct or complete the preceding problem items. In some cases, corrected items can also be incomplete.

corrected phrases: Phrases in which the corrected items occur are called corrected phrases. In most of the cases found in the corpus, corrected phrases are grammatically correct and complete phrases. Similar to corrected items, corrected phrases can be incomplete or interrupted, if the first repair attempt fails. That is to say, that corrected phrases can be disfluency-phrases as well.

disfluencies: Disfluencies are speech occurrences distinguishing written language from spoken language. They include simple phenomena such as silent pauses, filled pauses (interjections), repetitions, and restarts as well as more complex phenomena such as speech errors and speech repairs.

disfluency-phrase: Interrupted or incomplete phrases are called disfluency-phrases. They can also be phrases which are prosodically disfluent but grammatically complete. Problem phrases are in most cases disfluency-phrases, whereas corrected phrases are likely to be disfluency-phrases under certain circumstances.

disfluency structure: The internal structure of speech disfluencies is described by disfluency structure in terms of three components, namely the problem phrase (RP), the editing phase (EP) and the corrected phrase (CP). Based on these three elements, the majority of the speech disfluencies found in the entire corpus can be described.

edit signal hypothesis: Labov [57] suggested that a number of simple rules of interpretation are supposed to account for certain types of disfluencies. These rules are called edit signals in the literature. An edit signal hypothesis has also been applied by Hindle [42] in his deterministic parser developed for processing spontaneous speech input. Generally speaking, edit signals can be glottalization, interjections or silent pauses.

F_0 **reset:** A prosodic reset means that the prosodic properties of utterances do not follow the usual declination and somehow show similarity to certain previous locations. F_0 reset can be observed, when the pitch contour, represented by fundamental frequency values, is reset to that of the previous speech elements. This phenomenon can merely be observed, but not proved, because it is a null-hypothesis which can not be statistically proven.

internal disfluency relations: Internal disfluency relations are determined in terms of two linguistic characterizations: syntactic and prosodic features. In simple disfluencies, there is no need to consider the internal disfluency relation, because there is no contrast point, i.e. the corrected phrase, to reflect the relationship between the structure in the problem phrase and in the corrected phrase. From the syntactic aspect, the POS information in the initial elements of the problem and corrected phrases as well as the phrase boundaries between the problem and corrected phrases determining the structural form of disfluencies. From the prosodic aspect, three relations serve the same purpose, too. These are pitch, temporal and tonal pattern relations.

main interruption rule: Speakers frequently interrupt immediately after the reparandum; sometimes even interrupt the reparandum itself. Whether the phrase is completed does not seem to play an important role. This hypothesis is proposed by Levelt [66]. A large number of speech disfluencies found in the corpus which are interrupted within the reparandum itself empirically support this rule.

pauses: Pauses clearly indicate the occurrence of disfluent speech. They are divided into two sub-groups: filled and unfilled pauses. Filled pauses are called interjections, whereas unfilled pauses are simply silent pauses. In this thesis, the location and the duration of pauses are investigated. Usually, pauses are located between the problem phrase and the corrected phrase.

problem items: Problem items are words which can neither be found nor be accepted according to a language-specific lexicon and grammar. They can be speech errors or interruptions. Pauses and interjections are for instance problem items, as well as interrupted words or phrasal fragments which are not available in the lexicon or violate the syntactic rules in the grammar.

problem phrases: Phrases containing problem items are called problem phrases. Problem phrases are often disfluency-phrases which are incomplete phrasal fragments or disfluent speech strings.

prosodic markings of speech disfluencies: Prosodic markings of speech disfluencies refer to both perceptual and acoustic characteristics of prosodic features found within and around disfluencies in spoken language. This thesis focuses more on the production aspect of different realization forms of prosodic marking of speech disfluencies. These are markings of pitch contour, temporal features and tonal patterns.

reparandum: In the literature of speech repairs, the reparandum is the speech stretch containing inappropriate information or form. The reparandum is supposed to be repaired later in the flow of speech, however, it is not always the case that a repair takes place after the reparandum. In this thesis, the term *problem item* has the same function, but not the same content as the reparandum. The reason for giving up the term *reparandum* is that the domain of the reparandum has not been consistently defined in the literature.

repetitions: Repetitions are immediately repeated words. This thesis only accounts for direct word repetitions, not paraphrases of longer linguistic units such as clauses or sentences. Repetitions can be regarded as simple speech repairs, where the to-be-corrected elements are identical with the corrected elements.

restarts: Restarts occur, when the speaker stops his momentary planned utterance and begins with a new utterance without finishing the original utterance. In this case, both syntactic and prosodic disfluencies are often available.

speech errors: Speech errors are speech elements which are uttered improperly with respect to a lexicon and a grammar of a given language. The deviations can take place on the phonological, lexical, syntactic, morphosyntactic, and semantic levels. Speech errors are supposed to be repaired after they are made. However, not all speech errors were repaired in the disfluencies found in the corpus. Speech errors are classified in this thesis into the group of interrupted disfluencies.

speech repairs: Speech repairs include corrections and specifications of previously realized speech strings. Speech repairs are regarded as complex disfluencies where a repairing process takes place. Complex disfluencies of this kind have, as defined in the disfluency structure, values for the problem phrase and the corrected phrase instantiated.

Appendix B

Notations and Abbreviations

:p: a silent pause.

article$[X_1, ..., X_n]$**:** grammatical properties of the article. In German, X_n can be a sequence composed of following information:

grammatical gender:	*musculine, neuter, feminine*
numerus:	*singular, plural*
cases:	*nominative, possessive, dative, accusative*
definitiveness:	*indefinite, definite*

CP: corrected phrase

DFP: disfluency-phrase

DFS: disfluency structure

EP: editing phase

RP: problem phrase

Appendix C

Disfluencies

500 disfluencies which are identified by the author and are used for the corpus analysis are illustrated in the form **(problem phrase, editing phase, corrected phrase)**. In case of a second attempt to repair the problem phrase the *second corrected phrase* is given after each disfluency in - -.

0101 den den grünen quadratischen klotz

D011 die das rote

D012 ganz ganz

0102 der gel mhm der gelbe würfel

0103 ganz ganz

0103 zu anf anfang

0103 mit dem gelben k öh würfel

0104 im in der mitte

0105 die zweiten die zweite

0106 den prod propeller

0107 an dem an der g -an dem unteren-

D013 das das gewinde

0108 das blaue den blauen würfel

0201 zw

0202 auf den auf das stück

0203 die die schrauben

0204 die

0204 von von den sch -auf die schraube-

D021 auf der s

0205 von dem von dem

0206 fe

0207 eine eine sieben -eine siebener-

0208 o

0209 die die dreistange

0209 zwa

0209 ange öh befestigt

0210 von de

0210 die letzte das letzte loch

0211 an das an die verlängerung

0212 das der bug

0213 diese dieses farblose plastikzeug

0214 mach das gleiche machst

0215 vor an den gelben würfel

0301 in die in das

0302 so ein

0303 ein kleines auch ein kleines loch

0304 durch den durch dies rote -also durch das rad-

0305 diese vitt vi -violetten scheiben-

0306 die das

0401 ohne sche ohne kerbe

0402 von diesen von den orangenen gegenstück - mhmm diesen muttern-

0403 von den lange sch

0403 eine eine

0403 setzt uhm legst

0403 die den einen kurzen stab

0403 dass sich uhmm dass sie sich

0403 in auf der gleichen linie

0406 von dem von der kleinen scheibe

0407 ich jetzt habe ich

0408 auf das uhm auf -auf die öffnung-

D041 das das loch

0409 die wie den fünferstab

D042 den die gelbe schraube

0410 an dieses an diesen

0412 je wenn du jetzt

0412 das das

0413 auf der rech

0414 mit dem mit dem gelben

0415 ist ist

D043 mit der mit der kerbe

0416 dieses diese wei

0416 dieser weisse ring diese weisse scheibe

0501 die

0501 eine gel eine gelbe schraube

0501 mit äh eine runde

0501 eine runde schl - mit schlitz-

0501 eine auch eine rote

U 19 ein ähm ein würfel

0503 mit dem mit dem

0503 mit den drei mit den fünf löchern -mit dem langen stück-

0504 dies sch

0505 mit dem mit dem uh mittleren loch

0506 stehen steht

0506 dieses uhm fünf -diese schiene-

0507 ein eine

0508 blei

0508 mit einmal uhm mit der orangenen achkantschraube

0509 in das dri in das zweite

U 82 mit einer mit einem gewinde

0510 was was

U101 ein ein einzel -ein einzelstück-

U102 den roten einen roten würfel

U105 den dann leg den würfel

U106 die den arbeitsschritt

0511 auf auf das teil

U112 von von der seite

0512 zu dem zu der unteren schiene

U122 den den schwanz

0513 die das

0513 von der von dem

0514 im ersten a

0515 in in die

0515 in uh

U154 zu genau zu der schiene

0516 von dem m uhm

0517 mit der mit den orangenen schrauben

0518 die die runde

0518 was uh was

0519 mit einer mit einer mutter

0519 mit diesen mit dem einen

U189 der der

0521 sieht es autz aus

0522 die diese

U212 zwei zwei löcher

0523 auf das unbefe

0524 legst den da stellst

U246 eine öhm die rote schraube

0525 die schrau die rote schraube

0526 von diesen hölz hölzernen

0527 so dass es so dass es

0528 der der fehler

U286 es es

0529 da da

0530 der flü der flügel

U301 die erstmal die würfel

U302 das plätt

0531 das war uh das war

0532 befes

0533 einen einen

0534 auf den

0534 der der

U342 der die letzte abteilung

U354 bla

0535 da da

0535 die die lila -uhmm die lila ringe-

0536 auf die auf den untersten würfel

0536 sch uh schrauben

0536 sch schraubst

0537 geht ging

0602 mit s mit so einem ding

0603 nein nein

0603 nein nein

D061 mit ohne

0604 mit mit dem ding

0606 auf die auf das letzte loch

0607 rot rot

0607 qu

D062 den :p: äh

D063 von dem von der fünferplatte

0609 mit mit teil

0610 wo wo

0611 anstelle den äh des ersten würfels

0612 da du da du

0615 in richtung richtung

0617 diese die weissen räder

D064 auf :p:

0618 dieses äh dieses kleine weisse teil

D065 zeigt :p: zeigt

0619 muss muss

0620 ich ich

0621 dieser diese siebener

0622 das das gekreuzte teil

0623 so so

D066 unter dem :p:

0624 mit den mit den rädern

0625 nach o unten

0626 dem dem ende

0701 drehst drehst

D071 von den :p:

D072 hinter dem :p:

D073 die rund

D074 auf der lang

0702 so ein so ein -uhm so eine lange stange-

0702 diese die

0703 die or die gelbe schraube

0704 so so

D075 zu mit dem

0705 zur br blauen schraube

0707 von der

0707 in wenn du jetzt

0707 das das lange teil

0707 dieser dieses heck

0707 die rund uhm die eckige schraube

D076 das hinter hinterteil

0708 am ganz am ende

D077 hinter dem hinter dem grünen würfel

0710 dass die uhmm dass die blei -beiden uhmm-

D081 mit :p: mit eckig baufix

D082 mit :p: mit allen dreien

0801 wir nehmen wir

0801 die und den -dieser rote grosse würfel-

0802 die die

0802 die die -die rote schraube-

0803 die ro uhn orangenen mutter

0803 nach

0805 die ro gelbe runde schraube

0806 lich links

0807 die die andere fünfer

0808 das das

0810 zu dem zu der verlängerung

0811 die die muttern

0812 nein nein

0812 nein nein

0812 die gelbe mutter äh die orange mutter

0814 an den gr unter den grünen würfel

D083 der dieses flügels

D084 die :p: das loch

D085 mit den :p: zwei :p: zwei blauen mutter

D086 ha

D087 die :p:

0817 dieses gesch dieses rad

0817 nach der gelben mutter äh nach der gelben schraube

0818 die unt

0818 dieses die rote runde mutter

0819 mit dem :p:

0820 unter der der -der schraube-

0821 unter dem fünferdrei nee unter der dreierverlängerung

D088 kommt :p: kommt

0901 ist ist

0902 im auf die äusserste überlappungs -auf das äusserste überlappungsloch-

0903 scht mhmm

0904 mit ei -mit einem dreier-

0907 ganz ganz links

0907 überla lappt

D091 zwei ähm zwei löcher

D092 die die lorenz baufix

0908 von diesen diesen langen stangen

0909 für das für die andere gelbe schraube

D093 eine nee schr

D093 die die schlitzschraube

0910 die den -den zwei-

0910 von den überla beiden überlappungsteilen

0912 zwei mhm zwei stellen

0913 zusammensch

0914 mit dem schli schlitz

0915 ganz re ganz rechts

0917 die orange runde längs uh schritz -schlitzschraube-

D094 auf ähm auf

0918 das das

0919 das blaue äh also der blaue

0920 noch äh noch

0922 über uhmm überkreuzt

0923 das mhm das

0924 eine uh

0924 den letz den anderen siebener

0924 den grünen grünen würfel

0925 von oben von oben

0926 einen gelben den gelben würfel

0927 mit dem mit der grünen sechskantschraube

0927 oder oder

0927 dre mhm drehst

0928 gen blaues

0930 von der anderen seite von der anderen seite

0930 wenn wenn

D101 ein einen holzklotz

1003 sp

1004 mit mit fünf löchern

1006 auf auf dem dreilöchrigen stäbchen

1006 mit einem einer einkerbung

D102 das das rote kästchen

1007 die rote die rote -äh eckige schraube-

1008 dreilöchrigen :p: dreilöcheriges stäbchen

1010 vorne vorne

1011 nach nach dem

1012 jetzt jetzt

1013 die der -der propeller-

1014 mit mit uh drei löchern

1015 die die beiden

1016 durch die durch die mittlere loch

1017 ei :p:

1018 die die felge

1019 an

1019 mit mit einer einkerbung

1020 eine eine blaue eckige

1021 in der der mitte

D111 kleine äh diese kleinen :p: latten

1101 da da

1101 hint

1102 befestust befestigst

1103 in dieser blauen in diesem blauen dingen

D112 diese :p: diese zwei schienen

1104 musst uh musst

1105 an auf die -auf den blauen würfel-

1106 befestugst befestigst

1107 mit dem mit dem ersten

1108 die die längs

1109 auf auf die seite

1109 von dem sechseckigen von der sechseckigen schraube

1109 mit uh mit der gelben schraube

D113 in das über das loch

1110 in das in das mittlere loch

1111 auf der and

1112 von der dreischienen uh dreilöchrigen

1113 mit der mit den drei löchern

1114 die sch die andere gelbe schraube

1115 von der von der fünflöchrigen

D114 an an das andere ende

D115 zwischen zwischen

1116 mit der mittler mit dem mittleren loch

1117 die braunen uh die blauen langen schrauben

1118 dieses lilane diesen lilanen holzkreisel

1118 von von der anderen seite

1119 d

1120 die den grünen würfel

1121 mit der mittleren mit dem mittleren loch

1122 der rest hast den rest hast

D116 das das

1123 von der fünf fünflöchrigen

1124 dieses dieses -dieser rote würfel-

1201 wir wir

1201 einen

1201 so ein so eine art hubschraube

1202 so einen trag uh so eine kleine tragfläche

1203 dass dass

1203 von

1204 ein ein loch

1205 fol

1207 nur erstmal nur

1207 wir wir

1209 und und

D121 be beschriftet

D122 von dem von dem hubschrauber

1210 und ähmm und

1210 ist das ist

1211 wich

1212 die dreilis uh leiste

D123 jetzt jetzt

1213 mit den mit den muttern-mit den orangenen muttern-

1214 diese diese trap

D124 dann dann

1217 zu dem zu dem allerletzten löchel

1218 kommt kommt

1218 auf

1219 so was so was

1220 von unten von unten

1221 oben oben

1221 auf den -oben auf den roten würfel-

1221 mit der roten mit einer roten -mit der letzten roten schraube-

D125 in das das mittlere loch

1223 im re im rechten winkel

1223 zu der zu den uhmm leisten

D126 die die hinterste Tragfläche

1224 zu der zu -zu dem nächsten schritt-

1225 eine und eine mutter

1226 von ob

1226 die die flügel

1227 der erste flü -also der erste flügel-

1227 der die erste tragfläche

1228 was was

1228 und und

1228 für das für das -uh für das zweite loch-

1228 bei der bei der fünferleiste

1229 das ist das ist

1229 der der körper

D127 von vor

1230 das mindere mittlere lo -loch-

1230 das vordere lo loch

1231 also also

1231 mit mit der mutter

1232 dieses uh

1232 ich ich

1232 von der von dem -von dem-

1234 das das mittlere loch

D128 von dem von dem flugzeug

1236 genau genau

1236 unten und unten

1237 und und

1238 an das erste auf das erste loch

1238 von den

1238 das das loch

1238 ge ge-ähm geguckt-

1239 richtig richtig

1239 diese dieses loch

1240 zur fü früheren fünferleiste

D129 und unten :p: und unten

D12a einen einen blauen würfel

1241 unten unten -unten dran-

1242 das äh der

1242 es gibt es

1243 die flü uh die flügel

1243 wir jetzt gehen wir

1243 jetzt jetzt

1243 dre uh dreh

1244 dr zwei dreierleisten

1244 die die letzten

1246 in die in den blauen würfel

1246 vorne vorne

1246 in die mittelste in dies mittlere ding

1246 der der dreierleiste

D12b dann dann

1248 mit mit fehlern

1249 zwei zwei

1250 dass es dass es

1250 vier uh vier

D12c dies dieses teil

1251 zu den uh zu den rädern

1252 an grünen wird an grünen

1253 weil weil

1253 wir brauchen wir

1254 auf das flug

1254 auf das flug flugteil

D12d so ein so eine blaue schraube

D12e so ein plastik plastikteil

D12f ein orangen äh ein lila ring

D12g kommt dann kommt

1255 so dass so dass

1256 an die s an die seite

D12h erstmal erstmal

1301 das das ist das leitwerk

1302 ich den ich

1302 vor

1302 nein nein

D131 in das

1303 zum ro uh zum kopf

1306 an :p: an diesen dreierträger

D132 von

1307 mu sin musst

1401 die ro

1402 die diese eckige

1403 die orange äh die grüne

1403 die grüne grüne lange schraube

1404 die bleiden blauen schrauben

1405 dieses uh diese lila scheiben

D141 von :p:

1406 das das

1501 die b beiden kleineren

1501 mit so einer mit so einer rille

1502 hint

1503 mit so einen orangenen mit so einer orangenen schraube

1505 zur sei

D161 den den schwanz

1601 das das kleinere brettchen

1602 den schraubst das

1605 dieses violette diesen violetten ring

D171 ein rotes einen roten würfel

1701 eine sch gelbe schraube

1702 mit einem mit einer aufschrift

1703 die ge gelben

1704 eine lilane sch ein lilaner kreis

1802 sch

1802 diesen das längste

1802 das das

D181 auf das :p:

1805 v

1805 hinter den roten hinter die rote schraube

1806 so ein so ein eckiges

1806 ein eins -so eine mutter-

1807 gegenlehnen legen

D182 diese zwei orangenen äh violetten ringe

1808 dr uhm steckst

1809 mit der mit dieser verdickung

1809 so ein ora so ein violettes

1810 um umdrehen

1811 zu dr

1812 au auf die -an den gelben würfel-

1901 in die in das mittlere loch

1902 so ein dr so ein dreierscheibe

1902 so ein so ein platte

1903 dass das dass der

D191 das andere das andere loch

1904 eine pl uh -so eine dreierplatte-

1904 die uh die richtung

1904 die die platte

1904 längs

1905 auf das auf den reifen

1906 das b das blaue

1906 die die schraube

1907 an den an den würfel

D192 die die beiden übrigen

2001 einen einen dreier

2001 zum öh zur oberen leiste

2001 mit der letzten mit dem letzten loch

2002 auf diesen auf diese dreierleiste

2003 das das

2004 an dieser l

2005 am äh am ende

2005 die diese beiden langen siebenerleisten

2006 eine eine dreierlatte

2006 ein öh die beiden löcher

2007 an die an das vordere freie loch

2008 die uh die orangenen scheiben

2101 eine eine -ein teil-

2102 zwei zwei

2103 die die

2103 dass sich das so kreu dass es sich ein kreuz

2104 mit den

2104 die bei

2105 mit dr

2107 wo wo

2108 mittig mittig

2109 nimmst du jetzt legst du

2109 mit dem mit dem linken dreier

D211 die linke ha

2110 auf den auf den fünfer

2113 ist sind

2115 seit

2116 die die grüne -den grünen würfel-

2117 ni machst

2118 das w w

2201 nein nein

2201 mit einer mit der sechskantschraube

2203 von von der dreilöchrigen

D221 ange

2204 wievie

2205 mit mit einem endstück

2205 auf auf diese

2206 die die -die-

2207 so ein so ein blauen würfel

2208 nicht nicht

2209 die trag

Appendix D

Translation of Selected Disfluencies

speech repairs

```
(one ye one yellow bolt)
(one also one red)
(with the[neuter/masculine] :p: with the[neuter/masculine])
(with the three :p: with the five holes)
(with the[neuter/masculine] :p: with the[neuter] uh middle hole)
(one w)
(with once :p: uhm with the[feminine] orange eight-edge-bolt)
(in the[neuter] thr in the[neuter] :p: second)
(on on the[neuter] part)
(the[feminine] :p: the[neuter])
(from the[feminine] :p: from :p: the[neuter/masculine])
(in in the[feminine])
(with the[feminine] :p: with the[plural] orange bolts)
(the[feminine] the[feminine] round)
(with one[feminine] :p: with one[feminine] nut)
```

(with these :p: with the[neuter/masculine])

(the[feminine, plural] this[feminine] these uh plate)

(put it[masculine] there :p: lay)

(the[feminine] bo the[feminine] red bolt)

(so that it[neuter] :p: so that it[neuter])

(the[masculine] :p: the[masculine] mistake)

(it[neuter] was uh :p: it[neuter] was)

(the[feminine] the[feminine] purple)

(on the[feminine] :p: on the[feminine] lowest block)

(s uh screw)

(s :p: screw[2nd person, singular])

single word repetitions

(what what)

(what :p: uh what)

(there there)

(one[masculine] :p: one[masculine])

(the[masculine] :p: the[masculine])

(there :p: there)

Bibliography

[1] S. Abney. Parsing by Chunks. In Berwick, Abney, and Tenny, editors, *Principle-Based Parsing*. Kluwer Academic Publishers, 1991.

[2] L. Adriaens. *Ein Modell deutscher Intonation*. PhD thesis, Technische Universiteit Eindhoven, 1991.

[3] F. Althoff, G. Drexel, H. Lüngen, M. Pampel, and C. Schillo. The Treatment of Compounds in a Morphological Component for Speech Recognition. In *Natural Language Processing and Speech Technology. Results of the 3rd KONVENS conference*. Mouton de Gruyter, 1996.

[4] J. Bear, J. Dowding, and E. Shriberg. Integrating Multiple Knowledge Sources for Detection and Correction of Repairs in Human-Computer Dialog. In *Proc. of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 56–63, 1992.

[5] E. R. Blackmer and J. L. Mitton. Theories of Monitoring and the Timing of Repairs in Spontaneous Speech. *Cognition*, 39:173–194, 1991.

[6] D. Boomer and A. Dittman. Hesitation Pauses and Juncture Pauses in Speech. *Language and Speech*, 5:215–220, 1962.

[7] D. Brazil. Phonology: Intonation in Discourse. In van Dijk, editor, *Handbook of Discourse Analysis*, pages 57–75. Appleton Century Crofts, London, 1985.

[8] A. Butcher. Pause and Syntactic Structure. In Dechert and Raupach, editors, *Temporal Variables in Speech*, pages 85–90. Mouton, 1980.

[9] B. Butterworth, editor. *Language Production*, volume 1 of *Speech and Talk*. Academic Press, 1980.

[10] D. Byron and P. Heeman. Discourse marker use in task-oriented spoken dialog. Technical report, University of Rochester, 1998.

[11] J. Cahn. An Investigation into the Correlation of Cue Phrases, Unfilled Pauses and the Structuring of Spoken Discourse. In *Proc. of the IRCS Workshop on Prosody in Natural Speech*, pages 19–30, University of Pennsylvania Institute for Research in Cognitive Science, Philadelphia, 1992.

[12] H. Clark. Managing Problems in Speaking. *Speech Communication*, 15:243–250, 1994.

[13] H. Clark. *Using Language*. Cambridge University Press, 1997.

[14] W. E. Cooper and J. M. Sorensen. Fundamental Frequency Contours at Syntactic Boundaries. *Journal of the Acoustical Society of America*, 62(3):683–692, 1977.

[15] M. Coulthard. The Significance of Intonation in Discourse. In Coulthard, editor, *Advances in Spoken Discourse Analysis*. Routledge, 1992.

[16] A. Cutler, editor. *Slips of the Tongue and Language Production*. Linguistics – Special Issue: 19-7/8. Mouton, Berlin, 1982.

[17] N. Daly-Kelly. Linguistic and Acoustic Characteristics of Pause Intervals in Spontaneous Speech. In *ESCA EUROSPEECH '95*, pages 1023–1026, 1995.

[18] H. Dechert and M. Raupach, editors. *Temporal Variables in Speech*. Mouton, 1980.

[19] G. S. Dell. Representation of Serial Order in Speech: Evidence from the Repeated Phoneme Effect in Speech Errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10:222–233, 1984.

[20] G. S. Dell and P. A. Reich. Stages in Sentence Production: An Analysis of Speech Error Data. *Journal of Verbal Learning and Verbal Behavior*, 20:611–629, 1981.

[21] D. Duez. Perception of Silent Pauses in Continuous Speech. *Language and Speech*, 28:377–389, 1985.

[22] D. Duez. Acoustic Correlates of Subjective Pauses. *Journal of Psycholinguistic Research*, 22:21–39, 1993.

[23] P. Eisenberg. *Grundriß der deutsche Grammatik*. Verlag J.B. Meyer, 1994.

[24] R. Eklund and E. Shriberg. Crosslinguistic Disfluency Modeling: A Comparative Analysis of Swedish and American English Human-Human and Human-Machine Dialogues. In *ICSLP 98*, Sydney, 1998.

[25] U. Engel. *Deutsche Grammatik*. Julius Groos Verlag, 1991.

[26] R. Fiehler. Analyse- und Beschreibungskategorien für geschriebene und gesprochene Sprache. Alles eins? In Čmejrkova, Daneš, and Havlová, editors, *Writing vs Speaking. Language, Text, Discourse, Communication*, pages 175–180. Gunter Narr Verlag, 1994.

[27] K. Fischer and M. Johanntokrax. Ein linguistisches Merkmalsmodell für die Lexikalisierung von diskursgesteuerten Partikeln. Technical Report 18, SFB 360, 1995.

[28] C. Ford and S. Thompson. Interactional Units in Conversation: Syntactic, Intonational, and Pragmatic Resources for the Management of Turns. In Ochs, Schegloff, and Thompson, editors, *Interaction and Grammar*, pages 134–184. Cambridge University Press, 1996.

[29] V. Fromkin, editor. *Speech Errors as Linguistic Evidence*. Mouton, 1973.

[30] V. A. Fromkin, editor. *Errors in Linguistic Performance*. Academic Press, 1980.

[31] G. Gazdar, E. Klein, G. Pullum, and I. Sag. *Generalized Phrase Structure Grammar*. Harvard University Press, Oxford, 1985.

[32] D. Gibbon and C. Sassen. Prosody-Particles Pairs as Discourse Control Signs. In *EUROSPEECH 97*, 1997.

[33] J.J. Godfrey, E.C. Holliman, and McDaniel. SWITCHBOARD: Telephone Speech Corpus for Research and Development. In *Proc. of the IEEE Conference on Acoustics, Speech and Signal Processing*, pages 517–520, 1992.

[34] F. Goldman-Eisler. Pauses, Clauses and Sentences. *Language and Speech*, 15:103–113, 1972.

[35] F. Grosjean, L. Grosjean, and H. Lane. The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology*, 11:58–81, 1979.

[36] D. Gross, J. Allen, and D. Traum. The TRAINS 91 Dialogues. Technical Report 92-1, Dept. of Computer Science. University of Rochester, 1993.

[37] E. Gülich and T. Kotschi. Discourse Production in Oral Communication. In Quasthoff, editor, *Aspects of Oral Communication*, pages 30–66. de Gruyter, 1995.

[38] P. Heeman. *Speech Repairs, Intonational Boundaries and Discourse Markers: Modelling Speakers' Utterances in Spoken Dialog*. PhD thesis, University of Rochester, 1997.

[39] P. A. Heeman and J. Allen. Detecting and Correcting Speech Repairs. In *Proc. of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 295–302, 1994.

[40] P. A. Heeman and J. Allen. Tagging Speech Repairs. In *Proc. of the Human Language Technology Workshop*, pages 187–192, 1994.

[41] A. Hieke. A Content-Processing View of Hesitation. *Language and Speech*, 24(2):147–160, 1981.

[42] D. Hindle. Deterministic Parsing of Syntactic Non-fluencies. In *Annual Meeting of the Association for Computational Linguistics*, pages 123–128, Philadephia, 1983.

[43] J. Hirschberg and B. Grosz. Intonational Features of Local and Global Discourse Structure. In *Proc. of Speech and Natural language Workshop*, pages 441–446, 1992.

[44] J. Hirschberg and J. Pierrehumbert. The Intonational Structuring of Discourse. In *Proc. of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 136–144, 1986.

[45] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, languages, and Computation.* Addison-Wesley Publishing Company, 1979.

[46] D. James. Some Aspects of the Syntax and Semantics of Interjections. In *Papers from the Regional Meeting Chicago Linguistics Society*, volume 8, pages 162–172, 1972.

[47] D. James. Another Look at, say, Some Grammatical Constraints, on, oh, Interjections and Hesitations. In *Papers from the Regional Meeting Chicago Linguistics Society*, volume 9, pages 242–251, 1973.

[48] D. James. *The Syntax and Semantics of Some English Interjections.* PhD thesis, University of Michigan, Ann Arbor, 1974.

[49] P. Johnson-Laird and R. Byrne. *Deduction.* Erlbaum, 1991.

[50] B. Johnstone. An Introduction. Perspectives on Repetition. *Text*, 7(3):205–214, 1987.

[51] B. Johnstone, editor. *Repetition in Discourse Interdisciplinary Perspectives*, volume 1–2. Ablex Publishing Corporation, 1994.

[52] W. Kindt. Satzbegriff und gesprochene Sprache. *Lingua*, pages 25–48, 1994.

[53] B. Johnstone Koch. Repeating Yourself: Discourse Paraphrase and the Generation of Language. In *Proc. of the Eastern States Conference on Linguistics*, pages 250–259, 1984.

[54] K. Kohler. PROLAB- the Kiel System of Prosodic Labelling. In *Proc. ICPhS*, 1995.

[55] J.C. Kowtko and P.J. Price. Data Collection and Analysis in the Air Planning Domain. In *Proc. of the DARPA Speech and Natural Language Workshop*, pages 119–125, 1989.

[56] E. Kuipers. Ill-formed input and finite state devices. Master's thesis, University of Groningen, 1995.

[57] W. Labov. On the Grammaticality of Everyday Speech. In *Annual Meeting of the Linguistic Society of America*, volume 41, New York, 1966.

[58] W. Labov. *Sociolinguistic Patterns.* University of Pennsylvania Press, Philadephia, 1972.

[59] M. Lalljee and M. Cook. An Experimental Investigation of the Function of Filled Pauses in Speech. *Language and Speech*, 12:24–28, 1969.

[60] M. Lalljee and M. Cook. Filled Pauses and Floor-Holding: The Final Test? *Semiotica*, 12:219–225, 1974.

[61] B. Lang. A Generative View of Ill-Formed Input Processing. In *ATR Symposium on Basic Research for Telephone Interpretation*, Kyoto, 1989.

[62] H. Langer. Syntactic Normalization of Spontaneous Speech. In *COLING-90*, pages 180–183, 1990.

[63] J. D. Laver. The Detection and Correction of Slips of the Tongue. In Fromkin, editor, *Speech Errors as Linguistic Evidence*. Mouton, The Hague, 1973.

[64] I. Lehiste. Phonetic Disambiguation of Syntactic Ambiguity. *Glossa*, 7:107–121, 1973.

[65] I. Lehiste, J. P. Olive, and L. A. Streeter. The Role of Duration in Disambiguating Syntactically Ambiguous Sentence. *Journal of the Acoustical Society of America*, 60:1199–1202, 1976.

[66] W. J. Levelt. Monitoring and Self-Repair in Speech. *Cognition*, 14:41–104, 1983.

[67] W. J. Levelt. Spontaneous Self-Repairs in Speech: Processes and Representations. In Van den Broecke and Cohen, editors, *Proc. of the Tenth International Congress of Phonetic Sciences*, pages 105–118, Dordrecht, 1984. Foris.

[68] W. J. Levelt. *Speaking: From Intention to Articulation.* MIT Press, 1989.

[69] W. J. Levelt and A. Cutler. Prosodic marking in Speech Repair. *Journal of Semantics*, 2(2):205–217, 1983.

[70] R. J. Lickley and E. G. Bard. Processing Disfluent Speech: Recognising Disfluency Before Lexical Access. In *Proc. of the International Conference on Spoken Language Processing*, pages 935–938, Banff, 1992. ICSLP.

[71] R. J. Lickley, R. C. Shillcock, and E. G. Bard. Processing Disfluent Speech: How and When are Disfluencies Found? In *Proc. of the 2nd European Conference on Speech Communication and Technology*, pages 1499–1502, Italy, 1991. Eurospeech 91.

[72] R. J. Lickley, R. C. Shillcock, and E. G. Bard. Understanding Disfluent Speech: Is There an Editing Signal? In *Proc. of the International Congress of Phonetic Sciences*, pages 98–101, Aix-en-Provence, 1991.

[73] R.J. Lickley. *Detecting Disfluency in Spontaneous Speech.* PhD thesis, University of Edinburgh, 1994.

[74] R.J. Lickley and E.G. Bard. When Can Listeners Detect Disfluency in Spontaneous Speech? *Language and Speech*, 2(41):203–226, 1998.

[75] J. Local. Continuing and Restarting. In Auer and Di Luzio, editors, *The Contextualization of Language*, pages 273–296. John Benjamins Publishing Company, 1992.

[76] H. Lüngen, M. Pampel, G. Drexel, D. Gibbon, F. Althoff, and C. Schillo. Morphology and Speech Technology. In *Proc. of the ACL-SIGPHON Workshop on Computational Phonology in Speech Technology*, Santa Cruz, 1996.

[77] D. G. MacKay. The Structure of Words and Syllables: Evidence from errors in Speech. *Cognitive Psychology*, 3:210–227, 1972.

[78] MADCOW. Multi-Site Data Collection for a Spoken Language Corpus. In *Proc. of Speech and Natural Language Workshop*, pages 7–14, 1992.

[79] J. Martin. On Judging Pauses in Spontaneous Speech. *Journal of Verbal Learning and Verbal Behavior*, 9:75–78, 1970.

[80] R. Meringer and K. Mayer. *Versprechen und Verlesen: Eine Pschychologisch-Linguistische Studie.* Göschensche Verlagsbuchhandlung, Stuttgart, 1895.

[81] S. Nakajima and J. Allen. A Study on Prosody and Discourse Structure in Cooperative Dialogues. *Phonetica*, 50:197–210, 1993.

[82] C. Nakatani and J. Hirschberg. A Speech-First Model for Repair Detection and Correction. In *31st Annual Meeting of the Association for Computational Linguistics*, pages 46–53. Ohio State University, 1993.

[83] C. Nakatani and J. Hirschberg. A Speech-First Model for Repair Detection and Correction. In *Proc. of the ARPA Workshop on Human Language Technology*, pages 329–334, 1993.

[84] C. Nakatani and J. Hirschberg. A Corpus-Based Study of Repair Cues in Spontaneous Speech. *Journal of the Acoustical Society of America*, 95:1603–1616, 1994.

[85] C. Nakatani, J. Hirschberg, and B. Grosz. Discourse Structure in Spoken Language: Studies on Speech Corpora. In *AAAI-95 Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, 1995.

[86] E. Ochs. Planned and Unplanned Discourse. In Givon, editor, *Syntax and Semantics*, volume 12 of *Discourse and Syntax*, pages 51–80. Academic Press, 1979.

[87] D. O'Shaughnessy. Recognition of Hesitations in Spontaneous Speech. In *ICASSP '92*, pages 521–524, 1992.

[88] D. O'Shaughnessy. Locating Disfluencies in Spontaneous Speech: An Acoustical Analysis. In *Proc. of the 3rd European Conference on Speech Communication and Technology*, pages 2187–2190, 1993.

[89] J. Pierrehumbert. *The Phonology and Phonetics of English Intonation.* PhD thesis, MIT, 1980.

[90] C. Pollard and I. Sag. *Information-Based Syntax and Semantics.* CSLI, 1987.

[91] C. Pollard and I. Sag. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, 1994.

[92] A. Postma, H. Kolk, and D. Povel. Disfluencies as Resulting from Covert Self-Repairs Applied to Internal Speech Errors. In Peters, Hulstijn, and Starkweather, editors, *Speech Motor Control and Stuttering*, pages 141–147. Elsevier Science Publishers, 1991.

[93] P. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong. The Use of Prosody in Syntactic Disambiguation. *Journal of the Acoustical Society of America*, 6:2956–2970, 1991.

[94] R. Rath. "Satz" und "Äußerungseinheit". Syntaktische und interaktive Struktur in der Sprache? In Leupold and Petter, editors, *Interdisziplinäre Sprachforschung und Sprachlehre. Festschrift für Albert Raasch*, pages 197–216. Gunter Narr Verlag, 1990.

[95] M. Reyelt and A. Batliner. Ein Inventar prosodischer Etiketten für VERBMOBIL. Technical report, Verbmobil, 1994.

[96] H. Sacks, E. A. Schegloff, and G. Jefferson. A Simplest Systematics for the Organization for the Turn-taking for Conversation. *Language*, 50(4):696–735, 1974.

[97] G. Sagerer, H. Eikmeyer, and G. Rickheit. "Wir bauen jetzt ein Flugzeug": Konstruieren im Dialog. Arbeitsmaterialien. Technical report, SFB 360 "Situierte Künstliche Kommunikatoren". Universität Bielefeld, 1994.

[98] A. Sanderman. *Prosodic Phrasing. Production, Perception, Acceptability and Comprehension*. PhD thesis, Technische Universiteit Eindhoven, 1996.

[99] U. Schade. *Konnektionismus. Zur Modellierung der Sprachproduktion*. Westdeutscher Verlag, 1992.

[100] U. Schade and U. Laubenstein. Repairs in a Connectionist Language-Production Model. In K"ahler and Rieger, editors, *Contributions to Quantitative Linguistics*, pages 79–90. Kluwer Academic Publishers, 1993.

[101] E. Schegloff. Turn Organization: One Intersection of Grammar and Interaction. In Ochs, Schegloff, and Thompson, editors, *Interaction and Grammar*, pages 52–133. Cambridge University Press, 1996.

[102] E. A. Schegloff. The Relevance of Repair to Syntax-For-Conversation. In Givon, editor, *Syntax and Semantics*, volume 12 of *Discourse and Syntax*, pages 261–288. Academic Press, 1979.

[103] E. A. Schegloff. Discourse as an Interactional Achievement: Some Uses of 'Uh Huh' and Other Things That Come Between Sentences. In Tannen, editor, *Georgetown University Round Table on Languages and Linguistics 1981*, Analyzing Discourse: Text and Talk, pages 71–93, 1982.

[104] E. A. Schegloff, G. Jefferson, and H. Sacks. The Preference of Self-Correction in the Organization of Repair in Conversation. *Language*, 53(2):361–382, 1977.

[105] D. Schiffrin. *Discourse Markers*. Cambridge University Press, 1987.

[106] L. C. Schourup. *Common Discourse Particles in English Conversation*. PhD thesis, The Ohio State University, 1983.

[107] M. Selting. *Prosodie im Gespräch*. Max Niemeyer Verlag, 1995.

[108] S. Shattuck-Hufnagel. The Role of Word-Onset Consonants in Speech Production in Speech Production Planning: New Evidence from Speech Error Patterns. In Keller and Gopnik, editors, *Motor and Sensory Processes of Language*, pages 17–51. Lawrence Erlbaum, Hillsdale. NJ, 1987.

[109] S. B. Shimanoff and J. C. Brunak. Repairs in Planned and Unplanned Discourse. In Elinor, Ochs, and Keenan, editors, *Discourse Across Time and Space*, pages 123–167. Southern California Ocassional Papers in Linguistics, 1977.

[110] E. Shriberg. *Preliminaries to a Theory of Speech Disfluencies*. PhD thesis, University of California at Berkeley, 1994.

[111] E. Shriberg. Acoustic Properties of Disfluent Repetitions. In *Proc. of the International Congress of Phonetic Sciences*, pages 384–387, Stockholm, 1995.

[112] E. Shriberg, J. Bear, and J. Dowding. Automatic Detection and Correction of Repairs in Human-Computer Dialog. In *Proc. of Speech and Natural Language Workshop*, pages 419–424, 1992.

[113] E. Shriberg and R. J. Lickley. Intonation of Clause-Internal Filled Pauses. In *ICSLP,2*, pages 991–994, Banff, 1992.

[114] E. Shriberg and A. Stolcke. How Far Do Speakers Back Up In Repairs? A Quantitative Model. In *ICSLP 98*, Sydney, 1998.

[115] A. Stolcke, E. Shriberg, and R. Bates. Automatic Detection of Sentence Boundaries and Disfluencies Based on Recognized Words. In *ICSLP 98*, Sydney, 1998.

[116] M. Swerts and R. Geluykens. Prosody as a Marker of Information Flow in Spoken Discourse. *Language and Speech*, 37(1):21–43, 1994.

[117] M. Swerts, A. Wichmann, and R. Beun. Filled Pauses as Markers of Discourse Structure. In *Proc. of ICSLP 96*, pages 1033–1036, Philadephia, 1996.

[118] D. Tannen. Repetition in Conversation: Toward a Poetics of Talk. *Language*, 63(3):574–605, 1987.

[119] D. Tannen. *Talking Voices – Repetitions, Dialogues, and Imagery in Conversational Discourse.* Cambridge University Press, 1989.

[120] T. Taylor and D. Cameron. *Analysing Conversation.* Pergamon Press, 1987.

[121] J. Terken and R. Collier. Syntactic Influences on Prosody. In Tohkura, Vatikiotis-Bateson, and Sagisaka, editors, *Speech Perception, Production and Linguistic Structure*, pages 427–438. IOS press, 1992.

[122] D. Traum and P. Heeman. Utterance Unit in Spoken Dialogue. In Maier, Mast, and LuperFoy, editors, *Dialogue Processing in Spoken Language Systems*, Lecture Notes in Artificial Intelligence. Springer Verlag, 1997.

[123] J. Fox Tree. *Comprehension after Speech Disfluencies.* PhD thesis, Stanford University, 1993.

[124] J. Fox Tree. The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language*, 34:709–738, 1995.

[125] C. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. Price. Segmental Duration in the Vicinity of Prosodic Phrase Boundaries. *The Journal of the Acoustical Society of America*, 91(3):1707–1717, 1992.

[126] C. Van Wijk and G. Kempen. A dual system for producing self-repairs in spontaneous speech Evidence from experimentally elicited corrections. *Cognitive Psychology*, 19:403–440, 1987.