CHAPTER ONE

THE PRIMACY OF THE CITATION FORM
IN THE PHONOLOGICAL ANALYSIS OF
NON-CITATION FORM UTTERANCE

1.1 Introduction

In this chapter I will examine ways in which analysts have typically gone about accounting for certain phonetic aspects of language beyond the citation form, why they have adopted this approach, and why I feel there is a need for a different approach to the phonological description of language beyond the citation form. In 1.2 I will consider the types of data used, the type of description adopted, the special machinery proposed to cope with the phonetic shape of certain grammatical items in English and the reasons why the phonetics and phonology of the citation form have enjoyed analytical priority in the description of non-citation form utterance. In 1.3 I will argue against the use of the phonetic shape of the citation form as the analytical base for the description of non-citation form utterance, and will argue for the use of conversational material as a sole data source, a
requirement that entails the analysis of this data in its own terms without reference to phonetic descriptions of other linguistic activities.

1.2 Phonetic and phonological analyses of non-citation form utterance

In this section I will look at the ways in which analysts have gone about carrying out phonetic and phonological analyses of non-citation form utterance. I have split this review into four parts. In 1.2.1 I will examine the diversity of sources which analysts have used to provide data for their analyses, and the implications which the different sources have for the data and ultimately the analysis thereof. This diversity of sources and data has led me to adopt the negative term non-citation form utterance. In 1.2.2 I will examine the ways in which much of the phonetics observed in non-citation form utterance is accounted for. In 1.2.3 I will look at a particular area of research which is concerned with accounting for the phonetic patterns observable in certain grammatical items. I am referring here to work on clitics and cliticisation, which has led to the proposal of machinery to account for phonetic patterns in grammatical items which cannot be dealt with adequately using the types of phonological processes proposed to account for patterns observed in the lexis and other parts of the
grammar. In 1.2.4 I will ask why it is that the citation form has enjoyed analytical priority in the description of non-citation form utterance.

The works I will be concentrating on have appeared primarily over the last two decades, although I will also include a few works prior to this period to show that, even with the advent of many new phonological frameworks, the way of describing non-citation form utterance has changed very little.

1.2.1 The nature of the data

One of the most striking features of phonetic and phonological description of language beyond the citation form is the vast number of labels attached to the data being described. One finds 'relaxed' (Stampe 1973), 'sloppy' (Nespor & Vogel 1986), 'careless' (Sweet 1908), 'casual' (Zwicky 1972b, Stampe 1973, Hasegawa 1979, Lass 1984), 'informal' (Brown 1977), 'rapid familiar' (Gimson 1980), 'rapid colloquial' (Armstrong 1932, Ortiz Lira 1976), 'rapid casual' (Kreidler 1989), 'rapid' (Sweet 1877, 1900, 1908), 'fast' (Hasegawa 1979, Selkirk 1981, Nespor & Vogel 1986), and '(natural) connected' (Shockey 1974, Gimson 1980, Roach 1983, Hawkins 1984, Barry 1984). Barry (1984) says that 'the choice of one of these terms often reflects the assumption on the part of the writer as to what might be the principal cause of the variation in
question' (1984, 2). This is particularly true of Hasegawa (1979) who wants to distinguish 'fast' and 'casual' speech in Japanese on sociolinguistic grounds, the former being characterised by certain phonological processes acceptable in any interaction, the latter being restricted to informal interaction. Labels such as 'informal', '(rapid) colloquial/familiar' and 'casual' also reflect the social setting in which the language is being used. Sweet's (1908) use of 'careless', Stampe's (1973) 'relaxed' and Nespor & Vogel's (1986) 'sloppy' seem to have more to do with a particular mode of articulation than with social factors. On the other hand, the terms 'quick', 'fast' and 'rapid' reflect certain phonetic/phonological characteristics of fast tempo. However, in two of these works, Selkirk (1981) and Nespor & Vogel (1986), 'fast speech' is used to describe a type of talk which they are not examining, but which neatly accommodates those features which do not turn up in the type of non-citation form utterance they are examining. So, Nespor & Vogel (1986) assign 'certain cases of Vowel Deletion in American English' (1986, 23) to fast speech, 'resyllabification may be found across words... in fast or sloppy speech' (1986, 64) and 'restructuring might turn out to be more frequent in fast speech' (1986, 173). Likewise, Selkirk (1981) uses 'fast' as a convenient label for the place where patterns in the phonetics may be observed which are other than those already observed and accounted for in the data to hand. In this case it is
certain juncture phenomena:

'In colloquial or fast speech, the external sandhi rules may operate "more frequently".'
(Selkirk 1981, 20)

'... the [te] pronunciation before a vowel-initial verb feels extremely awkward, and is probably only possible in fast speech.'
(Selkirk 1981, 94)

'... a final consonant will not assimilate in voicing, except perhaps in very fast speech.'
(Selkirk 1981, 127)

Appealing to a type of non-citation form utterance which is essentially being described negatively gives language beyond the citation form a sense of 'anything goes'. However, I will argue, and the analysis I will provide in the empirical chapters of this study will show, that if one chooses to look at conversational material, the phonetic patterns are just as systematic as those to be found in citation form utterance.

What is often not indicated by the various labels attached to non-citation form utterance are the many different sources from which the data being analysed have been taken. The source of the data in turn has important implications for the data itself.

Analysts do not always state where they have taken the phonetic data they are analysing from, but even here it is often possible to identify the source on the basis of the type of data being presented.
I will now look at the three most commonly used data sources and show what implications these have for the nature of the data they provide.

Intuition is a great provider of data in many areas of linguistic research, and this is also the case in the provision of non-citation form utterance. Stampe's list of phonetic representations of the expression 'divinity fudge'\(^3\) (1973, 59) is the result of intuitions about well- and ill-formed pronunciations of the expression; Selkirk (1981) relies heavily on her own intuitions to provide her data, and at the same time she appeals to the reader's intuitions in support of the conclusions she has reached, e.g.

'I wager that speakers of English will find it much less easy to assimilate the final nasal of the underlined words below to the consonant that follows.'

(Selkirk 1981, 126)

Other cases of analysts using intuition to provide their data are less clear as they are not directly identified as such, but from the nature of the data being presented one can only assume that intuition has been used. Two examples of this are Kohler (1977) and Lass (1984). Kohler (1977), in a chapter entitled 'Segmentelle Satzphonetik', presents a number of phonetic transcrip-

---

\(^3\)I will consider Stampe's derivation of 'divinity fudge' in more detail in the next section.
tions representing different pronunciations of the same expression, "Hast du einen Moment Zeit". The transcriptions form a continuum from a phonetic representation of citation forms of the words in the sentence at one end to the 'stark verschliffene Aussprache' at the other:

(1)

1. [hast] [du:] [ʔainən] [mo' ment] [ˈtsait]
2. [ˈhas du (ʔ)ainən mo' men ˈtsait]
3. [ˈhas du ain mo' men ˈtsait]
4. [ˈhas ðə n mo' men ˈtsait]
5. [ˈhas ɾ n mo' men ˈtsait]
6. [ˈhas ɔː m mo' men ˈtsait]
7. [ˈhas (b) mo' men ˈtsait]

(adapted from Kohler 1977, 207-208)

In a similar fashion, Lass (1984) presents the different phonetic shapes of three noun phrases at different tempi:
My contention that Kohler and Lass have both used intuition is based on an elimination of what could not have provided such data. Neither set of data could have come from conversation. In Lass' case one could realistically expect the repetition of the same noun phrase a few times from recordings of a conversation. That such repetitions would be tokens of the same noun phrase at different tempi is highly improbable. Another possible source for this data would involve asking an informant to read the same expression a number of times, being asked to increase her/his delivery rate⁶. In order to see whether it is possible to elicit material anything like that which Lass and Kohler present I had a speaker of German utter "Hast du einen Moment Zeit" a number of times, asking her to utter quicker each time. As one might have expected the speaker's different versions of the expression exhibited considerable variation. However, this variation was a good deal different from the phonetics Kohler presents in two important respects. First, the differences across versions were not restricted to the pronoun-article

portion, but to the whole expression. Secondly, the phonetics of the versions I observed did not form a neat continuum of the type Kohler presents.

What remains is the analyst's description of what s/he considers to be possible versions of the same expression, i.e. the use of intuition. The provision of data by intuition is most commonplace in syntactic analysis. Although it is often claimed (e.g. Radford 1988) that linguists are primarily concerned with describing spoken language, it is clear from discussions about certain forms of ambiguity (e.g. 'I can not do it', 'the English teacher', etc.) that the intuitions sought are ones about the written language, and only at a later stage do observations about such features as stress in the spoken mode play a role. To what extent such intuitions are formed by the long years of instruction in using the written language at various educational establishments from an early age is a matter of speculation, but they must have some influence. The use of intuition to provide phonetic data, especially when this involves speculating about articulatory and phonatory activity occurring in talk beyond the citation form, can amount to little more than educated guesswork. There is no doubt that intuition is a very convenient way of collecting data, but there is no way of knowing that the data it throws up reflects what 'normal' speakers actually produce, and from my own observations of conversational material it is clear that intuition produces data which does not turn up in actuality, and at the
same time fails to provide material which does occur. Once one has cast doubt on the validity of data provided by intuition, as it does not seem to be representative of the data collected from speakers carrying out other linguistic activities such as text reading or conversation, one must also question the validity of any analysis based on such data.

Intuition, then, has been seen by some authorities to be a legitimate source of data; but it can be seen to have serious drawbacks.

The two other most commonly used data sources are (a) a type of controlled elicitation of material involving informants being asked to read or imitate text prepared by the analyst, and (b) the recording of conversation. Much work has involved the collection and comparison of data from these two activities. Shockey (1974) and Ortiz Lira (1976) both use data from text-reading and conversation. Nespor & Vogel (1986) use data from read sentences together with test sentences imitated by informants. The non-citation form utterance provided by conversation has been used as the sole data source by a few analysts\(^7\). In a number of works concentrating primarily on Stockport English, Lodge (1966, 1973, 1978, 1981, 1983, 1984) relies solely on data taken from conversations. Hurford (1968a

\(^7\)There has been a long tradition in work on intonation of using conversational material for at least part of the description, e.g. Jassem (1952), Crystal (1969), Halliday (1967 & 1970), Baldwin (1979).
& b, 1969 and 1970) also presents detailed impressionistic records of conversational material taken from Cockney English.

Data which the analyst is able to collect from these two activities is different in a number of respects. From the point of view of the analyst, text-reading has the advantage of allowing the analyst to control the elicitation process. The ability on the part of the analyst to control what a speaker produces in the course of a conversation is severely limited. Perhaps the best known example of an analyst controlling the elicitation in a dialogue situation is Labov (1972), in which the analyst used a question which would elicit a particular response ("fourth floor"). The data provided by reading aloud is influenced by such factors as the informant's prior experience of carrying out the task and the familiarity of the items in the text being read. Conversational material collected under circumstances where the subject is not aware that s/he is being recorded is the most natural in that it is least influenced, if at all, by the techniques which are used to collect it.

In this section I have surveyed a number of different works which all deal with non-citation form utterance. The analysts have provided a rich selection of labels which undoubtedly provide short descriptions of the data being handled, be it from an articulatory (e.g. 'fast', 'sloppy', 'rapid', etc.) or sociolinguistic (e.g. 'infor-
mal', 'casual', etc.) point of view. What I have also tried to show, however, is the diverse nature of the linguistic activities from which the data of non-citation form utterance is drawn, together with the implications which the data source can have for the data and ultimately for the analysis carried out on it (cf. analysis of data from intuition, p. 22). In section 1.3, where I will present arguments for a different approach to the analysis of non-citation form utterance, I will argue that only one of the three non-citation form data types listed can be considered a suitable counterpart to the citation form, namely conversation.

1.2.2 The nature of the description and of the described

Up until now I have been using the term 'non-citation form utterance' to cover a wide variety of other terms used by various authors to describe the material they are dealing with. However, 'non-citation form utterance' also rather neatly characterises the status of the data with respect to the analysis being carried out. What is common to all the works cited in the previous section (1.2.1) is the way in which the description of non-citation form utterance is carried out and the nature of the types of phenomena being described. The phonetic shape of an item in non-citation form utterance is derived, via a set of phonological processes, from the
phonetics or the phonology of the same item in its citation form. The phonetics of the citation form, then, provides the analytical base for the description of non-citation form utterance.

Not only is the way of describing non-citation form utterance uniform, there is also a striking similarity in the nature of the phonological processes which provide the derivation from the phonetics of the citation form to those of non-citation form utterance, i.e. the processes posited all have a very limited domain over which they operate: either segmentally (e.g. vowel deletion, flap deletion, shortening), or intersegmentally (e.g. assimilation, nasalisation).

The approach to the description and the nature of the processes posited are nearly identical across all forms of generative phonology. Stampe (1973), working in the framework of natural phonology presents part of the derivation of the expression 'divinity fudge'8:

\[(3)\]

<table>
<thead>
<tr>
<th></th>
<th>other processes:</th>
<th>*da.vínati</th>
<th>fádʒ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>syllabication:</td>
<td>*da.vín.a.ti</td>
<td>fádʒ</td>
</tr>
<tr>
<td>2.</td>
<td>flapping:</td>
<td>*da.vír.a.ti</td>
<td>fádʒ</td>
</tr>
<tr>
<td>3.</td>
<td>vowel-nasalisation:</td>
<td>da.vír.a.ti</td>
<td>fádʒ</td>
</tr>
<tr>
<td>4.</td>
<td>flap-deletion:</td>
<td>da.ví ě.a.ti</td>
<td>fádʒ</td>
</tr>
<tr>
<td>5.</td>
<td>syllabication:</td>
<td>*da.ví ě.ti</td>
<td>fádʒ</td>
</tr>
<tr>
<td>6.</td>
<td>vowel-nasalization:</td>
<td>da.ví ě.ti</td>
<td>fádʒ</td>
</tr>
<tr>
<td>7.</td>
<td>a-harmony:</td>
<td>da.ví ō.ti</td>
<td>fádʒ</td>
</tr>
<tr>
<td>8.</td>
<td>shortening:</td>
<td>da.ví.ti</td>
<td>fádʒ</td>
</tr>
<tr>
<td>9.</td>
<td>syllabication:</td>
<td>*da.ví.ti</td>
<td>fádʒ</td>
</tr>
</tbody>
</table>

8The strings preceded by an asterisk are those forms which Stampe considers to be unpronounceable.
10. flapping: \textipa{də.vɪˈɾ.i fʌdʒ} \\
11. flap-nasalization: \textipa{də.vɪˈɾ.i fʌdʒ} \\
12. flap-deletion: \textipa{də.vɪ.ɾ.i fʌdʒ} \\
13. syllabication: *\textipa{də.vɪˈɾ.i fʌdʒ} \\
14. vowel-nasalization: \textipa{də.vɪˈɾ.i fʌdʒ} \\

(Stampe 1973, 59)

This example shows rather well how a number of different processes are applied, beginning with one phonetic shape and gradually working down to a possible non-citation form token in 14. The processes have only one segment as the domain over which they operate and they reapply 'whenever the configurations they would eliminate arise' (Stampe 1973, 60). Within traditional generative phonology, Lass (1984) provides the derivation of 'hand-picked':

(4)
\begin{align*}
\text{Input} & : \text{hænd pɪkt} \\
\text{Length} & : \text{hæ:nd pɪkt} \\
\text{Stop-del} & : \text{hæ:n pɪkt} \\
\text{Assim} & : \text{hæ:m pɪkt}
\end{align*}

(Lass 1984, 300)

And in dependency phonology, Lodge (1984) accounts for the phonetic shape of items in his conversational material in terms of phonological processes such as various types of lenition\(^9\).

\(^9\)I will return to a more a detailed examination of Lodge's work, and in particular such processes as lenition and deletion, in Chapter Four.
The phonological processes themselves are stated in terms of context-sensitive rewrite rules, such as the following taken from Lass (1984) which provide the derivation of the phonetic shape [wɔʔn] ('one can'):

(5)

1 Vowel deletion

\[
\begin{align*}
[\text{V} &] \rightarrow \emptyset / [+\text{obs} \quad \text{N} \\
[-\text{stress}] &
\end{align*}
\]

2 Nasal syllabification

\[
\begin{align*}
[\text{N} &] \rightarrow [+\text{syl}] / [+\text{obs} \quad \text{N}] \rightarrow \{\#\} \\
[\alpha[\text{artic}] &] \rightarrow \{\#\} \{\text{C}\}
\end{align*}
\]

3 Nasalisation

\[
\begin{align*}
\text{V} \rightarrow [+\text{nas}] / \quad \text{N} \rightarrow \{\#\} \{\text{C}\}
\end{align*}
\]

4 Nasal deletion

\[
\begin{align*}
\text{N} \rightarrow \emptyset / [\text{V} \quad \text{N}] \\
[+\text{nas}] &
\end{align*}
\]

5 Glottal stop

\[
\begin{align*}
[+\text{obs} \quad \text{N}] \rightarrow \emptyset / ? \quad [\text{N} \quad [+\text{syl}]
\end{align*}
\]

(Lass 1984, 302-303)

At first sight these rules appear to be identical to the type of rewrite rules in citation form phonology which
derive a phonetic segment (allophone) from a phonological unit, such as a phoneme, e.g.

\( /p/ \rightarrow [\text{ŋp}] /\_
\)

A rule of this type is generally interpreted to the effect that the phonological unit /p/ is realised as a bilabial stop with simultaneous glottal closure when that phonological unit is at the final place in the phonological structure of the word. However, rules of the kind Lass presents are crucially different from allophonic rewrite rules in a number of respects. Instead of providing a phonetic realisation of a particular phonological unit at a stated place in the phonological structure of a word or syllable, these rules take the phonetic shape of a portion of utterance (segment) as their input and change it into another phonetic shape within the context of other adjacent portions of utterance. That such rules take phonetics as their input is clear from the tabular presentation of ever changing phonetic representations, with the rule accounting for each successive change standing at the side of the phonetic representation. Having identified that the rule input is phonetic\(^{10}\), as is the context which determines the change, it also becomes clear that many of the rules themselves are flawed, because they contain

\(^{10}\) So, for instance, in Lass' rules presented in (5) above, 'V' generalises over vocalic portions, e.g. [æ:] and does not generalise over a set of phonological units such as /i/, /ɛ/, /æ/, etc.
information which can only be phonological. In Lass' vowel deletion rule one stipulation on the input is that it carry the feature [-stress], a feature which can only be phonological having phonetic exponents such as pitch, loudness and vowel quality. Likewise, the rules of nasal syllabification and nasalisation rely on structural information in determining the correct context. The structural information I am referring to is $\#$, which symbolises the final place in the structure of the word. However, this sort of information is only available at the phonological level of abstraction and not in the phonetics, which is the only place where these rules can obtain their information.

Once one has identified flaws in the rules and the input to the rules, which can be generally characterised as a mixing of material from different levels of linguistic abstraction, i.e. the phonetic and the phonological, one starts to question the claims being made about the phonetic and phonological representations in certain analyses. So, for instance, Dressler (1972 & 1975) makes certain explicit claims about the input to his derivation of the phonetics representative of particular styles.
The table in (7) shows the different phonetic shapes of the prepositional phrase nach/zu Hause in three different varieties of German in different styles. At first sight it would appear that a clear distinction is being made between an abstract phonological representation (in slash brackets) of the phrase and the various phonetic output representations (in square brackets). It is not until one considers the rules\textsuperscript{1} which derive the 'phonetics' from the 'phonology' that one realises that both the input and output must be of exactly the same type, i.e. phonological, phonetic, or some mixture of the two. The

\textsuperscript{1}These are of exactly the same type as those given in (5). Many examples of the implementation of such rules on material from Breton are to be found in Dressler (1975) and Dressler & Hufgard (1980).
only way in which the input in (7) differs from the various outputs is that it looks different.

To summarise, the phonetic shape of a non-citation form item is derived via a set of processes, stated as a set of context-sensitive rewrite rules, from the phonetics of the same item in its citation form. The phonetics of the citation form are presumably (because this is never made explicit) derived from a set of 'conventional' allophonic rewrite rules. This can be represented schematically using Lass' (1984, 300) example of the derivation of 'hand-picked':

\[
\begin{align*}
(8) & \quad \text{Phonological representation} \\
& \quad \text{of citation form: } /\text{hænd pikt}/ \\
& \quad \text{Application of allophonic rewrite rules.} \\
& \quad \text{Phonetic representation} \\
& \quad \text{of citation form: } [\text{hænd pikt}] \\
& \quad \text{Application of 'phonetic' rewrite rules.} \\
& \quad \text{Phonetic representation} \\
& \quad \text{of non-citation form} \\
& \quad \text{utterance: } [\text{hæ:m pikt}] \\
\end{align*}
\]

In Chapter Four I will show that the flaws which these rules have are relatively insignificant when compared with their analytical value, i.e. whether they can
be seen to constitute a phonological analysis at all. Perhaps more importantly, the processes which have been proposed to change and delete various phonetic segments have become more than just a way of accounting for the phonetics of non-citation form utterance. These processes form a filter through which analysts look at their data, seeing the phonetics in terms of the processes they have already proposed. The consequence of this is that the analyst is prevented from providing any other account for the phonetic data before him/her. I will not provide a concrete example of this claim here as a good deal of Chapter Four is devoted to comparing the merits of two approaches to the same set of data, i.e. that described in this section and my own approach.

1.2.3 Accounting for the phonetic shape of grammatical items

The phonetic shape of certain grammatical items (determiners, modal and non-modal auxiliaries, prepositions, pronouns, etc.) in non-citation form English presents the analyst with a number of problems which do not occur when accounting for the phonetic shape of lexical items.

In the previous section I reproduced Stampe's (1973) derivation of 'divinity fudge'. Stampe found it sufficient to propose a number of segment domain processes
which together changed one string of phonetic segments into another. Certain grammatical items present other problems because they can exhibit very different phonetic shapes at different points in utterance. So, at certain points in utterance, the auxiliary 'have' can exhibit a phonetic shape with initial vocal cord narrowing, a relatively open front vocalic portion and a stricture of labiodental close approximation finally, and at other points it may have a shape which is no more than a relatively central vocalic portion. Not only are the phonetic shapes considerably different, but a particular phonetic shape of a grammatical item is restricted to certain points in utterance.

In earlier works (e.g. Soames 1899, Sweet 1900), and in introductory texts today (e.g. Gimson 1980, Roach 1983), the different shapes are dealt with in terms of lists of strong and weak forms, and they are described relatively informally. More formal attempts at accounting for the shape and distribution of grammatical items have been done in terms of clitics and cliticisation.

The category clitic arose to account for items in many languages which defy a satisfactory categorisation as syntactic words on the one hand, or morphological affixes on the other. Zwicky (1977) divided clitics into three groups: simple clitics, special clitics and bound words. All the English weak forms fall into the first of these groups, i.e. simple clitics. The shapes of certain grammatical items in English are categorised as clitics.
for the following reasons\textsuperscript{12}:

(i) like affixes, they cannot stand in isolation, e.g. the various weak forms of the auxiliary 'have', orthographically 've, 'd, 's \textsuperscript{13}.

(ii) like affixes, they are structurally dependent on adjacent words for their phonetic shape, e.g. the third person singular form of 'have', 's, is subject to the same allomorphy as the plural affix in the nominal system and the third person singular affix in the verbal system.

(iii) unlike affixes, they are not sensitive to the syntactic category of the item they attach to, e.g.

a) The person I was talking to's died. PREP.

b) The boy you hit's crying. VERB

c) Anyone who's naughty's going to be shot. ADJ.

d) Driving quickly's stupid. ADV.

(iv) unlike affixes, they can be related to a strong form, i.e. a word, having similar semantic content and syntactic distribution.

The ways in which the weak form of a grammatical item is arrived at (cliticisation) differ with respect to the places in the grammar where cliticisation is carried out. In many works (Akmajian & Heny 1975, Selkirk 1984,

\textsuperscript{12}This selection is adapted from Zwicky (1977, 1984, 1987) and from Zwicky & Pullum (1983).

\textsuperscript{13}I will return to the significance of clitic representation later in this section.
Kaisse 1985) the derivation of a weak form begins in the syntactic component of the grammar with the structural binding of a grammatical item onto an adjacent word via Chomsky adjunction. This then creates the appropriate structure which can either feed phonological processes, such as destressing and vowel reduction in the case of Selkirk (1984), or, in Kaisse (1985), create the environment for a particular weak form allomorph from the lexicon. Berendsen (1986) approaches the derivation slightly differently. Although he follows Kaisse (1985) and has the weak form allomorphs of grammatical items listed in the lexicon alongside their strong form counterparts, the binding of a clitic to its host is carried out in phonological terms by constructing metrical trees which rely on syntactic structure. The tree structures then provide appropriate environments for the operation of phonological processes of the type I presented in the previous section.

In this section I have presented reasons why some writers have found it necessary to construct special machinery to account for the phonetic shapes of certain grammatical items, machinery different from that needed to provide the correct phonetic shapes of lexical items and other grammatical items. Rather than arriving at the right shape via phonological processes alone, the derivation of a weak form begins in the syntax, and in certain works (Kaisse 1985, Berendsen 1986, Zwicky 1987\textsuperscript{14}) the weak

\textsuperscript{14}Berendsen and Zwicky both follow Kaisse in this respect.
forms of certain auxiliaries in English are listed separately in the lexicon.

What I shall do now is to ask whether this machinery is necessary, and ultimately whether one needs the category clitic at all. For the time being I will only address certain issues regarding the phonetic and phonological representation of clitics as a good deal of the work contained in the empirical chapters of this study deal with a phonological analysis of grammatical fragments of English.

One thing at least is uncontroversial: certain grammatical items exhibit very different phonetic shapes at various points in non-citation form utterance. The issue is how to account for these different shapes. The arguments I presented above (p. 34) justifying the category clitic and the arguments which Kaisse (1985) presents for listing weak form allomorphs in the lexicon can be shown to have a number of weaknesses.

First, it is claimed that the weak forms of grammatical items cannot stand alone. There are at least two linguistic levels where one can make such a claim:

(a) in the phonetics.

(b) in the phonology.

As all the works I have cited represent clitics phonetically, i.e. in square brackets, the claim is that the phonetic shape cannot stand alone, and this is where the 'isolation' argument crumbles. It is quite easy to
demonstrate that the phonetic shape of any item, as it is found in the context of an utterance, cannot stand alone. A very clear example of this, taken from Suffolk English is [pəʊə] (put it). Leaving aside for the moment the question as to where all the phonetics of 'put' reside in this representation, it is impossible for [pəʊə] to stand alone in this variety of English. If the 'isolation' claim is made at the phonological level (in the more informal accounts of grammatical items (e.g. Gimson 1980) the strong and weak forms are represented phonologically) then the issue becomes one of how to represent grammatical items phonologically in the first place. As I will show in later chapters, given a sufficiently abstract phonological representation of a grammatical item one can arrive at its phonetic shape at a particular point in utterance by appealing to its place in the syntactic, rhythmic and interactional structure.

Another reason for proposing the machinery of clitics and cliticisation in English is that the phonetic shapes of grammatical items end up 'looking' and behaving like certain nominal and verbal affixes, and the phonetic shape of a grammatical item at one point in utterance weak form is considerably different from the phonetic shape of the same item at another point. As I said above this considerable difference in phonetic shapes of certain auxiliaries in English leads Kaisse (1985) to propose the different phonetic shapes be listed separately in the lexicon. Three of Kaisse's arguments in favour of taking
this line are that reduced forms of many auxiliaries are subject to deletions not productive elsewhere in English (e.g. w-deletion in would and will, and h-deletion in has, have and had.), that these deletions are not even regular (e.g. there are no forms of was and were with deleted w), and that the type of vowel reductions observable in auxiliaries (e.g. what's, I've) happen at rates of speech where other items do not allow such reductions. However, the peculiarities which Kaisse identifies arise as a result of adopting a phonological analysis which is monosystemic. So, for instance, the labial velar approximant (Kaisse's w) one can observe in the phonetic shape of will and would at certain points in utterance is treated as being phonologically the same as the labial velar approximant found in the phonetic shape of was and were, and presumably the same as that in the nouns will and wood. The insistence upon a monosystemic phonology inevitably leads Kaisse (and others) to account for the different phonetic shapes of will and would at other levels of linguistic abstraction, i.e. lexical and syntactic. If one chooses instead to adopt a phonological account which is polysystemic, then the apparent peculiarities which Kaisse identifies disappear, because under such an analysis one would expect, among other things, that the phonological statement needed to account for the auxiliary system, or even certain items in that system, would be different from that needed for, e.g. the lexical class of nouns. Indeed, auxiliaries exhibit many other
peculiarities which Kaisse ignores or fails to recognise\textsuperscript{15}, e.g. at certain places in the word and syllable structure of lexical items, for instance, one has to set up a three way nasal system of bilabiality, apicality and dorsality. At the final place in the structure of auxiliaries this three way contrast is absent, and it is only necessary to state that nasality contrasts with check for instance\textsuperscript{16}.

In summary, accounting for the phonetic shape of certain grammatical items, here the clitic, has led to the construction of machinery different from that needed to account for the shape of lexical items. But I have outlined flaws in the arguments proposed to support the category clitic, and have given some indication of the ways in which the 'peculiarities' in the phonetic shapes of grammatical items can be overcome by adopting a phonological analysis which is polysystemic.

1.2.4 The reasons for the primacy of the citation form in the analysis of non-citation form utterance

In 1.2.2 I have shown how phonological descriptions of non-citation form utterance have been taken by some

\textsuperscript{15}For many other apparent 'oddities' in grammatical items, see Simpson (to appear).

\textsuperscript{16}For an account in these terms of the phonology of BE + ING in orthodox and aberrant speech, see Kelly & Local (1989, 190-202).
authorities to be dependent upon the phonetic shape of citation form utterance. In this section I will consider why the citation form has been used as the analytical base. Indeed, in the majority of the works on non-citation form utterance this does not seem to be a question worth answering: it is just assumed that this is the path to take. There are a few authors, however, who do consider in some detail why the citation form is fundamental to their analyses.

Kohler (1977) provides three reasons for using 'die phonetische Form des Einzelwortes' as a basis for changes observable in words in sentential environments. I will consider each of these reasons separately.

'Die Ausgangsformen sind, zunächst als theoretische Konstrukte, derart angesetzt, daß sich eine Vielfalt von disparaten Erscheinungen unter wenigen Aussagen fassen läßt, was nicht möglich wäre, wenn sofort der Satz, und nicht das Einzelwort den Ausgangspunkt bildete.' (Kohler 1977,211)

Here the phonetics of the citation forms are treated as theoretical constructs which stand in a one-to-many relationship with the many and varied phonetic shapes of the same items in their non-citation form. The task of constructing a set of phonological statements to account for the different phonetics found in non-citation form utterance is considered to be impractical, if not impossible, if the phonetics of the citation form are not there at the outset.
In Kohler's next reason for adopting the phonetics of the citation form as the analytical base, the citation form is no longer purely a theoretical construct:

'Wörter und mit ihnen die isolierte Aussprache stellen für die Sprecher sicher in irgend einer Form eine psychologische Realität dar und sind wohl neurophysiologisch im Kortex festgelegt.'

(Kohler 1977, 212)

Here the citation form has achieved psychological reality, albeit tentative. One could equally claim, at our present level of understanding about the workings of the brain, that the citation form does not represent a psychological reality: there is no way of supporting or refuting Kohler's claim. Indeed, recent work in other areas of linguistics (e.g. Gazdar et al. 1985, 5) has been careful to state that the theory of syntax being presented is not a psychological or biological theory of language, and that it is not the job of the linguist to concern her/himself with matters of psychology.

Kohler's third reason for taking the isolated word form as a basis is a purely practical one:

'Das Zugrundelegen des Einzelwortes hat auch einen praktischen Wert, daß die Erstellung von Aussprache-wörterbüchern in kanonisierter kontextfreier phonetischer Form erlaubt, die in einem Einleitungsteil oder in einem separaten phonetischen Handbuch wie den vorliegenden durch Regeln der Satzphonetik ergänzt werden können.'

(Kohler 1977, 212)

Here the phonetics of the citation form are seen in
their practical application in language learning and the like. Again, it is questionable whether this practical value is of any significance in work of a purely descriptive nature. It is worth making a comparison here between works on syntactic theory (e.g. Gazdar et al. 1985, Radford 1988) and pedagogical grammars of languages (e.g. Swan 1988, Leech & Svartvik 1975). Both types of work describe and provide an account of similar data, but their aims and applications are different. A syntactician would not have one eye on teaching application while proposing a set of rules to account for the set of combinatorial possibilities of words in a language.

Although Kohler proposes three different reasons for adopting the phonetics of the citation form as an analytical base, I consider only the first of these to be of immediate concern, i.e. the contention that citation form phonetics are a necessary theoretical construct without which one could not account for the many different shapes of non-citation form items.

Linell (1979) argues several points for what he calls 'the primacy of careful pronunciations':

'(a) It seems to fit intuition. When speakers are asked to give the pronunciation of a word form, they tend to give it in its most articulated form, since giving a reduced rendition would leave out some part of the knowledge they have.
(b) One can derive reduced variants from the maximally segmentalised variant, but not always vice versa, since some articulatory reduction rules are neutralising.
(c) In perception, there seems to be a common
process of reinterpreting more structure into the perceived phonetic signals than is actually present in reduced speech... Likewise, speakers seem to believe that they articulate more carefully than they actually do.

(d) People's intuitions about what are phonologically permissible, i.e. pronounceable, forms in the language are clearly based upon careful pronunciations.

(e) There is also behavioral evidence that conditions on careful pronunciations sometimes have interesting consequences for the nature of reduction, whereas conditions on some level of reduced pronunciations have no such status.

(f) In practice most traditional, structuralist and generative discussions of the phonology of languages have, often implicitly, been restricted by and large to careful pronunciations, which indicates that linguists have taken these to be the phonetic data of the language.'

(Linell 1979, 56-57)

At least three of the above - (a), (c) and (d) - rely on the intuitions of either the analyst, or the speaker, and I have already made clear my reservations about the use of intuition to provide data in section 1.2.1.

A further strand which is present in a number of arguments appears, initially at least, to be relatively unproblematic. Linell refers to careful pronunciations as being the 'most articulated form' in (a) and the 'maximally segmentalized variant' in (b). At first sight, it might seem to be rather a simple task to identify that the phonetic make-up of an item at one place, e.g. in its citation form, is more articulated or has more segments
than the same item in the environment of a sentence\textsuperscript{18}. A good example of this are the clitics presented in 1.2.4. An aspectual auxiliary such as 'has' pronounced in isolation is likely to have an open glottis at its onset, a relatively open and front period of vocalicness extending through the first portion, followed by a period of close approximation at the alveolar ridge. The same auxiliary in a sentential environment might just have a period of alveolar close approximation as its phonetics. From the articulatory point of view the phonetic shape of isolated 'has' involves both more movements of the vocal organs and takes longer to produce. However, at the final place in the syntactic and interactional structure 'has' might well have more phonetics than it has when read off a word list. I have observed this 'moreness' to involve at least two features. The duration of vocalic portion can be considerably longer than that associated with the vocalic portion of the item in isolation. One can also find changes in the quality of the vocalic portion, involving a relatively central off-glode before the onset of the alveolar approximation. On the basis of this, I would argue that Linell is wrong in claiming that the 'careful

\textsuperscript{18}There is no simple way of quantifying the 'degree of segmentalisation' and in my arguments here I am making use of what I consider to be a highly controversial type of 'segment counting' found in works such as Ortiz Lira (1976). This procedure seems to involve counting the number of segments present in the impressionistic records which the analyst has made of an utterance. This is very dubious exercise since it involves counting symbols which represent time-sliced portions of articulatory and phona-
tory activity in a highly arbitrary fashion.
pronunciation' is the 'most articulated' or 'maximally segmentalized' version of an item.

Linell's fifth argument, (e), is also subject to similar criticism. This argument is based on Gårding's (1974) study of consonant cluster reduction in Swedish:

'... syllable-final consonants are transposed to the next syllable if and only if thereby a syllable-initial cluster is created which would be permissible also on the level of careful pronunciations... Such clusters then turn out to be immune to reduction\(^9\), while other types of clusters are obligatorily or optionally reduced. I interpret this preservation of syllable-initial clusters as an interesting piece of evidence that at least some of the structures defined by constraints on careful pronunciations have a substantial significance also for speech behaviour at other phonostylistic levels.'

(Linell 1979, 121)

I do not want to deny that the observations made here reflect the phonotactic constraints set up on the basis of the type of careful pronunciations which Linell refers to, but rather to question whether such careful pronunciations are the only place where one could observe the patterns which would lead one to set up such phonotactic constraints in the first place. For, as I showed above, Linell's careful pronunciations are not the only place where lexical and grammatical items have 'a lot of phonetics'.

Linell's final argument, (f), is one of analytical convenience and a suitable paraphrase might be: "This is

\(^9\)emphasis Linell's.
the way it has always been done, so why change now?"

Lass (1984), in his chapter on 'Non-static phonology', has this to say about the material contained in his book up to that point:

'In keeping with phonological tradition we have had little to say about phonological structure as it is deployed in the normal contexts of language use, or with the properties of connected speech'

(Lass 1984, 294)

This is an uncontroversial statement about the analytical procedures used in phonology and phonetics both today and in the past. However he goes on to say this:

'But the simplified version of structure we've been using does have a special analytical priority; the complicated and often messy things we consider here require it as a basis, and are largely incomprehensible without it. Variation and instability, the "dynamic" aspects of phonology, are uninterpretable save against an invariant, stable background.'

(Lass 1984, 294-5)

This statement is more open to argument. Lass does not explain what he means by 'complicated' and 'messy'. For something to be 'messy', there must be something else which is considered to be 'unmessy'; this is presumably the 'simplified version of structure'. Therefore these are subjective judgements made by the analyst comparing the 'known' with the 'unknown', and not a more general statement about the nature of language itself. This statement
has in common with Kohler's first argument that it considers any analysis of non-citation form phonetics to be impossible without the presence of citation phonetics as the starting point.

In this section I have examined the arguments presented by three analysts for the need to have the phonetics of the citation form as the analytical base. Many of these arguments rely heavily on analyst/speaker intuitions. From other arguments it is clear that the citation form phonetics represent a highly convenient, if not necessary, starting point from which to analyse non-citation form utterance.

1.3 Arguments for a different approach

In this section I will argue for the need to approach the phonological analysis of non-citation form utterance differently. In 1.3.1 I will show why I consider the citation form to be an unsuitable analytical base. In 1.3.2 I will argue for using conversational material as the only legitimate source of material and why I consider it necessary to attempt to analyse this data in its own terms without reference to data from elsewhere.
1.3.1 Arguments against the citation form

In 1.2.4 I looked at the reasons why the citation form has been taken as the basis for the analysis of non-citation form utterance. I will now present a number of reasons why I consider the citation form to be an unsuitable starting point.

My first argument is in part a continuation of the arguments presented in 1.2.4 against Linell's 'maximally segmentalized' and 'most articulated' forms. If one asks a speaker to read a word from a word-list, it is possible that the phonetics of the word in isolation are those that the speaker might produce for the same item at a particular place in, say, conversation. A concrete example of this from Tyneside English illustrates this point very well. Local et al. (1986) give as one of the phonetic features attendant upon a certain class of turn-endings (cf. 1984, 416) the marked aspirated release of final voiceless plosives, and it is precisely this feature which one also observes in informants reading word-lists. An analysis of this variety which were to take the phonetic shape of the citation form as its base would be obliged to delete this aspirated release in those places in non-citation form utterance where it is absent. And the feature of aspirated release itself would not be attrib-

\footnote{The word-list behaviour was related to me in a personal communication from John Local.}
uted any status in conversation since in those places where it was present, it would be no different from the citation form.

One far-reaching implication of observations such as those described above is that one could argue that one cannot know what articulatory and phonatory aspects of an item produced in isolation are attributable to the phonology of that item without the analysis of the same item in its conversational context. This, of course, reverses Lass' (1984) argument regarding interpretability (see p. 46 above), as it is his 'clean' data which only becomes fully interpretable when one has looked at the 'mess', i.e. conversational material.

Secondly, if one takes the citation form as the basis for analysing conversational data, then the spoken material is drawn (using different means and strategies) from speakers carrying out radically different linguistic and interactional tasks. The phonetics of the citation form are drawn from the extreme end of spoken prose; whereas the data of conversation is taken from the use of language in the interaction of two or more speakers in natural surroundings.

Thirdly, the pronunciation of a word in isolation may often be 'artificial'. I am not exactly clear what 'artificial' means in this respect, although it is a term used by Gimson (1980, 263) and I shall assume it to be an adequate description of the pressures put on any speaker placed in the position of having to read a word list in

49
the environment of a studio, together with the sociolinguistic effects of spelling pronunciation, speaking 'clearly', etc.

My fourth argument against the use of the citation form is connected with the first. If one chooses to describe non-citation form utterance from a variety of English other than the standard (which I shall), then one might well be confronted with a set of citation form utterances whose phonetics have more in common with the phonetic exponents of the phonology of the standard variety. It is precisely this that sociolinguistic studies of phonological variables have investigated. The observations of similarities and differences between the phonetics of citation and non-citation form utterance in a non-standard variety, together with the proposal of phonological rules and structures to account for these observations would therefore be of sociolinguistic importance, but could not be seen as providing an adequate descriptive account of the phonology of the non-standard variety itself.

Finally, it seems to me that the whole enterprise of deriving the phonetics of non-citation form utterance from those of the citation form is 'back to front'. Work on animal behaviour, e.g. Strohecker & Tschanz (1991a & b), comparing the behaviour of certain animals in the wild with behaviour in a domestic environment, makes observa-

---

21e.g. Labov (1972), Dressler et al. (1972), Trudgill (1974).
tions about the similarities and differences in the behaviour in the different environments, and is able to cast light on certain aspects of the domestic behaviour on the basis of observations made in the wild. Nowhere is an attempt made to derive the behaviour in the wild from that in the domestic environment. But it does seem that phoneticians/phonologists are using an approach very similar to this when accounting for data taken from the most natural form of linguistic behaviour, i.e. conversation, in terms of the data taken from the highly artificial situation of a speaker reading a word-list.

1.3.2 Conversational data as the analytical base

Having proposed a number of reasons why one should not take the phonetic shape of the citation form as the analytical base, I shall now present the arguments for using conversation as the only legitimate data source and analysing this data in its own terms without reference to data collected from other linguistic activities. To begin with I will provide a definition of what I consider conversation to be.

'Conversation' is a term which, like so many other terms in linguistics, has been defined in various ways. It is not possible, at the present stage of research, to describe conversation in linguistic terms. Perhaps the most spurious use of the term is to be found
in texts designed for the language learner. Here, the acquisition of conversational skills implies the ability to recognise and use certain features of the connected speech of spoken prose which are deemed necessary to attain fluency in a foreign language. So Scott's book entitled 'English Conversations' (1942) is a phonetic reader designed for the foreign learner of English. The texts consist of invented dialogues transcribed systematically, marking stress, together with other features such as syllabic consonants and the weak forms of grammatical items. The texts do not represent impressionistic phonetic records of talk between two or more people. Instead they are transcriptions of the spoken prose used by the author himself.

As there is no linguistic definition of conversation, one must look to particular interactional activities carried out by speakers, and it is in these terms that linguists have defined conversation. Abercrombie (1965a) defines conversation in the following way:

'Under "conversation" I would include all those linguistic occasions when there is the opportunity for give and take; when it is understood, that, at least in theory, there is more than one active participant, however long one of the participants may go on for.'

(Abercrombie 1965a, 2)

\[^{22}\text{emphasis Abercrombie's.}\]
Conversation here would presumably include the interaction of doctor-patient, shopkeeper-customer, together with parent-child, and adult-adult. Stubbs (1983) points to two features:

'...(a) language which occurs naturally without any intervention from the linguist;... (c) language which is spontaneous in the sense of unplanned, and which is composed in real time in response to immediate situational demands;...'
(Stubbs 1983, 33-34)

These features are opposed to those describing the language of intuition and spoken prose. However, neither of these narrows down conversation to the extent that they exclude activities such as talking to oneself, or the situation where a person is called to give an impromptu speech, where the hearers are passive participants in the talk.

Levinson (1983) gives an initial definition of conversation as:

'that familiar predominant kind of talk in which two or more participants freely alternate in speaking, which generally occurs outside specific institutional settings like religious services, law courts, classrooms and the like.'  (Levinson 1983, 284)

Later on he gives a 'more technical characterisation of what conversation is':

'Conversation as a unit... is characterisable in terms of overall organisations of the sort sketched here in addition to the use of conversational
activities like turn-taking...

(Levinson 1983, 318)

This is a definition of conversation made on the basis of analyses of others who have assumed that the data they are analysing is conversation, and it is a characterisation of this sort which I shall use. The crucial feature which sets conversation apart from other sorts of linguistic activity is the interactional activity of turn-taking. Turn-taking is Abercrombie's (1965a, 2) 'opportunity for give and take', and it is not included in other sorts of linguistic activity such as lectures. Plays, although those taking part are seen to be 'taking turns' at speaking, are better seen to be reciting chunks of spoken prose in alternation.

There are a number of reasons why I am going to use conversation as the sole data source and why I want to analyse the conversational data in its own terms.

Perhaps the most compelling reason of all is that conversation represents 'the most natural, the most frequent, and the most widespread occurrences of spoken language' (Abercrombie 1965a, 3).

Secondly, although conversation is often associated with labels such as 'complicated' and 'messy' (cf. Lass 1984), work which has been done on conversation not within the field of phonetics and phonology\(^{23}\) has shown quite

\(^{23}\)The reader is referred to Taylor and Cameron (1987) for a recent survey of work which has been carried out on conversation over the past six decades.

54
convincingly that on an interactional level conversation is well organised and very systematic. The little work which has been done on the phonetic and phonological aspects of turn-taking (Local & Kelly 1986, Local et al. 1985 & 1986, Kelly & Local 1989, 263-286) shows that there are systematic patterns in the articulatory and phonatory complexes which participants in talk use and attend to in the course of carrying out various interactional tasks, such as turn completion and turn holding. One would expect from this that conversational data used to construct phonological statements about syllable structure, etc., would also exhibit regular, albeit complex, patterns and would lead to more well-founded generalisations about phonological entities.

Finally, contrary to Shockey's claim that 'the phonological differences between reading and conversational speech are more quantitative than qualitative' (Shockey 1974, 66), the empirical chapters of this study will show that any attempt to account for conversational material with reference to other more 'primitive' data, such as citation form utterance, can only lead to one missing generalisations.