Assessing Listening

Gary Buck
## Contents

*Series Editors’ Preface*  
*Acknowledgements*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 An overview of listening comprehension</td>
<td>1</td>
</tr>
<tr>
<td>2 What is unique to listening</td>
<td>31</td>
</tr>
<tr>
<td>3 Approaches to assessing listening</td>
<td>61</td>
</tr>
<tr>
<td>4 Defining the construct</td>
<td>94</td>
</tr>
<tr>
<td>5 Creating tasks</td>
<td>116</td>
</tr>
<tr>
<td>6 Providing suitable texts</td>
<td>154</td>
</tr>
<tr>
<td>7 Designing and constructing complete assessments</td>
<td>194</td>
</tr>
<tr>
<td>8 Illustrative tests of listening comprehension</td>
<td>209</td>
</tr>
<tr>
<td>9 Summary and future developments</td>
<td>247</td>
</tr>
</tbody>
</table>

*Bibliography*  
*Index*  

*Page x*

*xii*

*Page 1*

*Page 31*

*Page 61*

*Page 94*

*Page 116*

*Page 154*

*Page 194*

*Page 209*

*Page 247*

*Page 258*

*Page 271*
CHAPTER ONE

An overview of listening comprehension

Introduction

Listening comprehension is a process, a very complex process, and if we want to measure it, we must first understand how that process works. An understanding of what we are trying to measure is the starting point for test construction. The thing we are trying to measure is called a construct, and our test will be useful and valid only if it measures the right construct. Thus, the first task of the test developer is to understand the construct, and then, secondly, to make a test that somehow measures that construct. This is construct validity, and ensuring that the right construct is being measured is the central issue in all assessment.

The purpose of this book is to look at the listening process, and to consider how that should be measured. In this chapter, I will begin by examining the listening process, and how language is used to convey meaning. Much of this relates to reading as well as listening. Then, in Chapter 2, I will discuss what is unique to listening comprehension.

Different types of knowledge used in listening

If we consider how the language comprehension system works, it is obvious that a number of different types of knowledge are involved: both linguistic knowledge and non-linguistic knowledge. Linguistic knowledge is of different types, but among the most important are
phonology, lexis, syntax, semantics and discourse structure. The non-linguistic knowledge used in comprehension is knowledge about the topic, about the context, and general knowledge about the world and how it works.

There has been much debate about how this knowledge is applied to the incoming sound, but the two most important views are: the **bottom-up** view, and the **top-down** view. These terms refer to the order in which the different types of knowledge are applied during comprehension.

The **bottom-up vs. top-down views**

It is my experience that when people start thinking about language processing, they often assume that the process takes place in a definite order, starting with the lowest level of detail and moving up to the highest level. So they often assume that the acoustic input is first decoded into **phonemes** (the smallest sound segments that can carry meaning), and then this is used to identify individual words. Then processing continues on to the next higher stage, the syntactic level, followed by an analysis of the semantic content to arrive at a literal understanding of the basic linguistic meaning. Finally, the listener interprets that literal meaning in terms of the communicative situation to understand what the speaker means. This is the bottom-up view, which sees language comprehension as a process of passing through a number of consecutive stages, or levels, and the output of each stage becomes the input for the next higher stage. It is, as it were, a one-way street.

However, there are some serious problems with this view of language comprehension, and both research and daily experience indicate that the processing of the different types of knowledge does not occur in a fixed sequence, but rather, that different types of processing may occur simultaneously, or in any convenient order. Thus, syntactic knowledge might be used to help identify a word, ideas about the topic of conversation might influence processing of the syntax, or knowledge of the context will help interpret the meaning.

It is quite possible to understand the meaning of a word before decoding its sound, because we have many different types of knowledge, including knowledge of the world around us. In most situations we know what normally happens, and so we have expectations about
what we will hear. These may be very precise, or very vague, but while we are listening, we almost always have some hypotheses about what is likely to come next. In such cases it is not necessary to utilise all the information available to us – we can just take in enough to confirm or reject our hypotheses. To take a well known example, if we hear the following uncompleted sentence, ‘she was so angry, she picked up the gun, aimed and ______’ (adapted from Grosjean, 1980), we know what is going to happen, and we probably need very little acoustic information to understand the final word, be it ‘fired’, ‘shot’ or whatever. As we listen, we will expect a word such as fired, and we will probably process only enough of the sound to confirm our expectations, or we may not even bother to listen to the last word at all. Our background knowledge about guns and what angry people do with them helps us to determine what the word is. This is a top-down process. Similarly, when we part from a friend, we hear the words of parting, not so much by processing the sound of what she says, but because she is waving to us and saying something as she leaves. We need very little acoustic information to determine whether she is saying ‘good-bye’, ‘see you later’, or whatever, and we may not even bother to find out.

Listening comprehension is a top-down process in the sense that the various types of knowledge involved in understanding language are not applied in any fixed order – they can be used in any order, or even simultaneously, and they are all capable of interacting and influencing each other. This is sometimes referred to as an interactive process, especially by reading theorists.

However, we should not underestimate the importance of the acoustic input, nor the importance of the linguistic information. The point is simply that listening comprehension is the result of an interaction between a number of information sources, which include the acoustic input, different types of linguistic knowledge, details of the context, and general world knowledge, and so forth, and listeners use whatever information they have available, or whatever information seems relevant to help them interpret what the speaker is saying.

In the rest of this chapter we will look at each type of knowledge used in understanding spoken language. The chapter will be organised into five main sections:

i the input to the listener;

ii applying knowledge of the language;
using world knowledge;
iv the context of communication;
v building mental representations of meaning.

The input to the listener

Listeners listen to spoken language, and this is very different from written language. There are three characteristics of speech that are very particularly important in the listening comprehension construct: firstly, speech is encoded in the form of sound; secondly, it is linear and takes place in real time, with no chance of review; and thirdly, it is linguistically different from written language. In this section we will consider each of these characteristics in detail.

The acoustic input

The external input into the listening comprehension process is an acoustic signal. This represents the meaningful sounds of the language, the phonemes. These phonemes combine together to make up individual words, phrases, etc. This acoustic signal is often very indistinct; in normal speech, speakers modify the sounds considerably and not all the phonemes are clearly and unambiguously encoded in the message.

Phonological modification

The process by which sounds are modified is regular and rule governed. It depends on a set of phonological rules which vary from one language to another. In normal-speed speech, some sounds are modified by the sounds next to them; some are simply dropped, others are combined in complex ways. For example, many words that are quite clearly intelligible within a text are difficult to recognise in isolation (Pickett and Pollack, 1963; Pollack and Pickett, 1963). In other cases words, or word clusters, which we assume to have different pronunciations often prove to be very difficult to distinguish in isolation. For example, the difference between the two sentences 'I
wish she would’ and ‘I wish you would’ is usually minimal, and most speakers could not even tell the difference between their own utterances if they heard them out of context. The fact is that many words are quite indistinct, and it is the surrounding context that enables us to identify them with little trouble.

**Stress and intonation**

However, while there is ample evidence that many of the individual sounds may be either indistinct or missing, there is also plenty of evidence to suggest that, at least in English, this is not the case with the prosodic features of the language, the stress and intonation. This remains important even in very fast speech. The English stress pattern gives each word an individual form, which is as much a part of the sound of the word as the actual phonemes. Furthermore, speakers stress what they think is important, and the most important words, those that express the core meaning, get additional stress (Brown, 1990). Similarly, the intonation pattern of the utterance is usually very important. In English, intonation patterns are closely related to the structure and meaning of the text. For example, intonation indicates clausal boundaries, marks questions, and also indicates when it is appropriate for the listener to respond (Cooper, 1976; Garro and Parker, 1982). One of the most important aspects of listening comprehension is paying attention to stress and intonation patterns.

**Redundancy and shared knowledge**

Language is by its nature extremely redundant (Cohen, 1975, 1977), and there are so many clues to what the speaker is saying that listeners can understand, even if speakers do not speak very clearly. Speakers know this and instinctively modify their speech depending on the situation, and their knowledge of the listener. Generally, people who share knowledge of a topic will tend to speak faster, run the words together more and be far less distinct when speaking to each other. But when they are speaking to someone who has less background knowledge, they will tend to speak more slowly and with much clearer enunciation (Hunnicutt, 1985). Thus, we find words with a high information content, that is non-redundant words, to be
more clearly articulated than redundant words which have a low information content (Liebermann, 1963; Oakeshott-Taylor, 1977).

Given the fact that the acoustic signal is often indistinct, we might well ask how comprehension takes place. The answer is simple: we use our knowledge of the language to ‘replace’ any missing information. And this is where redundancy comes in – because language is redundant, we do not need all the information to be clearly expressed, we only need enough to activate our knowledge, we can then construct the meaning ourselves.

The real-time nature of spoken language

Speech takes place in real time, in the sense that the text is heard only once, and then it is gone. We cannot go back to a piece of speech and hear it again (although modern recording technology does actually allow this, most conversations take place without being recorded). Of course, we can often ask a speaker to repeat what they said, but strangely, speakers virtually never do. We almost never get the same words, but a re-statement in a different way: speakers realise that there is a problem, and usually try to help by re-phrasing or by offering examples. And even if we do get the same words, the stress and intonation are different for repeated information. In normal language use, we get just one chance at comprehension, and only one. There are two consequences of this. The first is that the listener must process the text at a speed determined by the speaker, which is generally quite fast. The second is that the listener cannot refer back to the text – all that remains is a memory, an often imperfect memory, of what was heard.

The necessity of automatic processing

Speakers generally speak very quickly: three words a second is quite normal. This leaves little time to think about the precise meaning of each word, or the way relative clauses are structured, or to speculate on what the pronouns might refer to. The words are flying past very quickly, and in order to understand speakers at this speed, the listening processes must be almost entirely automatic.
It is helpful to make a distinction between two types of cognitive process: **controlled processes**, which involve a sequence of cognitive activities under active control and to which we must pay attention; and **automatic processes**, which are a sequence of cognitive activities that occur automatically, without the necessity of active control and usually without conscious attention (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). This distinction is perhaps best illustrated in activities like learning to drive a car: at first the whole process is controlled and we have to pay conscious attention to everything we do, but after a while things become a little more automatic and we start doing things without having to think about them very much, until eventually the whole process becomes so automatic that we may not think about it at all. The difference between controlled and automatic processing is very important in second-language use. When second-language learners learn some new element of a language, at first they have to pay conscious attention and think about it; that takes time, and their use of it is slow. But as the new element becomes more familiar, they process it faster, with less thought, until eventually the processing of that element becomes completely automatic.

Given the speed and complexity of normal speech, the more automatic the listener’s processing, the more efficient it will be, and the faster it can be done; and conversely, the less automatic the processing, the more time will be required. For language learners with less automatic processing, comprehension will suffer. As the speech rate gets faster, they will not have sufficient time to process everything, so they will start paying proportionally more attention to lexical and grammatical processing and less attention to the wider interpretation of the meaning (Lynch, 1998). Then, as the speech rate gets even faster, the listener will have insufficient time to process even the lexical and grammatical information, and they will begin to miss parts of the text. At a certain speed, their processing will tend to break down completely, and they will fail to understand much at all.

This rarely causes a problem for first-language listeners, but the normal state with many second-language listeners is that the language is only partly known, and so language processing will be only partly automatic. In such cases processing will periodically break down because the listener cannot process the text fast enough.
There are many reasons why the listening process may go wrong. This could be due to background noise, or listeners may have their attention distracted, or be thinking of something else. Second-language listeners could have other difficulties: unknown vocabulary, complex syntax, or the text could be just too fast, for example. In all these cases, when listeners try to recall the content of the text, their representation of what the text was about is likely to be incomplete – the listeners’ interpretations will be inadequate, and will obviously vary.

Interpretations also vary even when the listening does not go wrong, and the whole issue of what listeners should understand from a text is very complex. Different listeners often understand different things from the same text (Brown, 1995a). This may be due to the effects of background knowledge. When we listen we use our background knowledge of the world to set up expectations, and then use those expectations to help us comprehend what we hear. If the topic of the text accords well with the listener’s world knowledge, then it will be much easier to understand than a text with a topic that the listener knows nothing about (Spilich et al., 1979; Pearson and Johnson, 1978). So, a talk on a subject about which the listener knows nothing will be more difficult to understand, and a talk which in some way violates or contradicts the listener’s expectations will be even more difficult to understand, and could cause considerable confusion even though the language may not be linguistically challenging. It is difficult to assimilate and remember something that does not seem to make sense.

Different listeners often have different motives for listening, due to different interests and different needs. Listeners will pay more attention to those features of a text which they think are more interesting or more relevant. Thus what listeners get out of a text will depend on the purpose for listening as well as their background knowledge, and interpretations will therefore often differ from listener to listener. There is no such thing as the correct interpretation of many spoken texts, only a number of different reasonable interpretations (Brown and Yule, 1983).

While it is very true that interpretations reasonably vary, a note of caution is in order. Competent listeners will usually all grasp the same information from explicit statements, such as announcements, and they will usually share much common gist after hearing a piece of
spoken discourse. If this were not the case, communication would be very difficult or even impossible.

**Linguistic features of spoken texts**

Most people assume that the language of speech is much the same as the language of writing. Well, this is not true. Speech and writing are both variants of the same linguistic system, but there are some considerable differences between them. Good descriptions of spoken language can be found in Carter and McCarthy (1997) and McCarthy (1998).

One important point, for example, is that people do not usually speak in sentences. Rather, spoken language, especially in informal situations, consists of short phrases or clauses, called idea units, strung together in a rather loose way, often connected more by the coherence of the ideas than by any formal grammatical relationship. The vocabulary and the grammar also tend to be far more colloquial and much less formal. There are many words and expressions that are only used in speech, never in writing.

**Planned and unplanned discourse**

The real-time nature of spoken language not only affects the listener, but also affects the speaker, who must speak in real time. This means that speakers must construct the text at a rapid rate, and attempt to organise and control the flow of information with little preparation time. Consequently most spoken texts are just a rough first draft. This is usually referred to as unplanned discourse (Ochs, 1979) – it is spontaneous and produced without taking much time for planning and organisation. Planned discourse may be thought of as polished, worked text. Often there is a considerable difference between planned and unplanned discourse. We think of something and then say it almost immediately; and what we produce, and what our listeners have to listen to, will consist of initial ideas, and first reactions, loosely or poorly organised, fragments of language, with hesitations, false starts, restatements, vocabulary repair, and even grammatically 'incorrect sentences' (Hatch, 1983).

The grammar of unplanned spoken discourse tends to be different from planned discourse. Normally in planned discourse the relation-
ship between the ideas or **propositions** (a proposition is a short statement that says something about something, i.e. one simple little fact) is expressed by means of the syntax. However, in unplanned discourse the context itself can be used to connect the propositions: they may simply be placed next to each other, or strung together, so that the listener has to make the right connections between the ideas in the text. Similarly, referents are often missing, and the listener may have to infer who, or what, the speaker is talking about.

**Linguistic differences between speech and writing**

Spoken idea units usually contain about as much information as we can comfortably hold in working memory, usually about two seconds, or about seven words (Chafe, 1985). They may consist of only one proposition, but usually they will have a number of propositions. In English, each idea unit usually has a single, coherent intonation contour, ending in a clause-final intonation; it is often preceded and followed by some kind of pause or hesitation. Idea units are often clauses insofar as they contain a verb phrase along with noun phrases and prepositional phrases. However, some idea units do not have a verb, or the verb is understood but not explicitly stated. Other languages have idea units that are structured differently.

Although idea units are a characteristic of spoken language, Chafe also claims that they can be recognised in written texts. Probably the nearest written equivalent to spoken idea units is not the sentence, but the **t-unit**, which is one main clause plus all its dependent clauses. If we regard these as the written equivalent of idea units, we can then use these as a basis for examining the major linguistic differences between spoken and written language (Chafe, 1985):

- In spoken language idea units tend to be shorter, with simpler syntax, whereas written idea units tend to be more dense, often using complex syntax, such as dependent and subordinate clauses, to convey more information.
- In spoken language idea units tend to be strung together by coordinating conjunctions (**and**, **or**, **but** etc.), whereas written idea units tend to be joined in more complex ways.
- Spoken language usually has hesitations: pauses, fillers and repetitions that give the speaker more thinking time, as well as repairs,
such as false starts, corrections in grammar or vocabulary, and afterthoughts.

- Spoken language tends to have more non-standard features such as dialect, slang, and colloquialisms, whereas written language tends to be far more formal, conservative and ‘correct’.
- Spoken language tends to be far more personal, with more emotional involvement and much less precision. Speakers tend to indicate their feelings more, with expressions such as ‘I think’ or ‘I mean’, or by making direct reference to the listener. They also tend to be content with gross approximations, or use overstatements and exaggerations.

These differences were developed through study of English, and I am not aware of any research which addresses the extent to which they are universal. However, the differences between speech and writing seem likely to be important in most, if not all, languages, although the extent and significance of the differences may vary.

Such differences are of course really a question of degree, and Tannen has argued that texts can be ranged along an oral–literate continuum, with oral texts at one end, having more characteristics of spoken language that are typically associated with casual conversation, and literate texts at the other end, having more characteristics of written language that are especially associated with expository written prose (Tannen, 1982, 1985).

This idea of a continuum of features is very useful for our purpose. The features that determine where texts go on the continuum include linguistic features (the sound system, phonology, stress, intonation), paralinguistic features (tone of voice, gestures and facial expressions), vocabulary choice, grammar, degree of planning, the type and degree of shared context between the participants, and the function of the text. It is important to realise that spoken texts will differ in their degree of orality depending on the listening situation, and an important aspect of listening test construction is to identify those features of oral texts that need to be included on our tests.

The listening situation

The situation in which the listening takes place can have a considerable effect on various aspects of the listening process. Firstly, the
situation can determine the topic. Talk in a bread shop is likely to be about bread, and in a chemistry lecture, about chemistry. Similarly, a chat with a friend will require listening to informal language, and a public speech will usually require listening to more formal language.

The degree of interaction between the listener and speaker

One of the most important aspects of the listening situation is the degree of interaction between the listener and the speaker. In some situations, the listener’s role may be non-collaborative – requiring nothing more than interpreting the speaker’s utterance. But in other situations, the listener’s role may require collaborative listening – making appropriate requests for clarification, back-channelling, making responses to interactional language, or taking responsibility for organising turn-taking. This can be regarded as a continuum. At one end is the situation where there is no possibility of an interaction: listening to the radio, television or a lecture are examples of this. Other situations, such as classrooms or presentations, are mainly one-way, but there is some chance to ask clarification questions, or to intervene. Further along the continuum would be a situation where a group of people talk together, with one person doing most of the talking, and the others listening and saying very little, but able to intervene if they wished. At the other end of the continuum, there is a truly interactive conversation between two people, who collaborate equally to maintain the conversation, taking turns to speak with equal rights to participate.

The degree of interaction is related to the extent that the listener must collaborate with the speaker to maintain the conversation. In a typical two-way conversation, the listener and speaker change roles, back and forth, and they collaborate together to manage the conversation. In such conversations, decisions about who talks, and when, are not random, but are determined by a very complex set of shared, turn-taking rules (Rost, 1990). These rules depend on the relationship between the two interlocutors, the topics expected to be discussed, the context, and so on. Some conversations follow prescribed formulas, such as ordering a meal in a restaurant, and others have particular purposes which often determine interlocutor roles, such as a consultation with a doctor. But in other situations, turn-taking is negotiated between the participants.
Speakers generally use intonation and other markers to indicate when they want to pass on the turn, and listeners often indicate by verbal and non-verbal means when they would like to take a turn. And as topics shift in conversation, one contributor, and then the other, will take control of the conversation. Most groups have unspoken rules that determine how this should be done, and the listener must understand these and respond as appropriate.

The listener’s responsibility to respond

In many situations, the listener also has responsibilities to respond in predetermined ways. For example, in the USA it is more or less obligatory to respond to ‘thank you’ with ‘you’re welcome’, and there are lots of other situations where some reasonably fixed response is expected: greetings, inquiries about health, compliments, and so on. If the listener fails to respond in the appropriate way, the interaction may not go well.

Another responsibility of the listener is to provide back-channeling. All the time we are listening we must give signs to the speaker that we are understanding what is being said, and that we are paying attention. This takes the form of formulaic expressions such as ‘really’, ‘yeah’, ‘oh yes’, ‘I see’, ‘how interesting’, ‘well I never’, and so forth. It will usually be accompanied by appropriate body language and other non-verbal signals. Back-channelling is very important, and if the listener ever stops providing this, the speaker will quickly sense that something is wrong, and will likely stop speaking.

The function of the interaction

We also need to consider the function of the interaction. We can make a distinction between two types of language use: transactional language, where the primary purpose is to communicate information, and interactional language, where the primary purpose is social interaction (Brown and Yule, 1983). Transactional language has a definite purpose, usually to get something done; it is important that the message be understood, and it matters that the listener gets it right. Examples are a teacher giving a homework task, a customer...
making a complaint, a doctor giving instructions to a patient, or two colleagues planning their work schedule.

Interactional language is quite different. Here the purpose is to maintain social relationships. What is important is not the content of what is said, but the fact that something is said. This includes greetings, comments about the weather, what is happening in the world and other phatic conversation. Generally the participants do all they can to make their conversation agreeable to each other: they express agreement, regardless of their real opinions, and pretend to understand, whether they do or not. In most listening situations, there is both transactional and interactional language use, although one will usually be more dominant in any particular situation.

Clearly, there are certain listening situations that make demands on the listener that are not normally thought of as listening skills: for example requests for clarification, taking responsibility for turn-taking, back-channelling, and making the appropriate responses to interactional language.

Applying knowledge of the language

In order to process language, we need to have knowledge of the language and the ability to apply that knowledge. It is useful to make a distinction between two types of knowledge: declarative knowledge and procedural knowledge. Declarative knowledge is the knowledge of facts or about things; procedural knowledge is knowledge about how to do things (Anderson, 1976). For example, knowing the fact that English relative clauses are introduced by who, which or that is declarative knowledge, but being able to string them together with other words to make relative clauses is procedural knowledge. In language use, declarative knowledge is of very limited value; in practical terms, something is not really known until it can be used, correctly and efficiently. This means that it is procedural knowledge which is important for listening performance, and as language testers this is what we should be mainly concerned with.

This section is divided into three parts: the first part deals with understanding individual words, the second part with understanding and processing sentences or idea units, and the third part with understanding longer discourse.
Understanding words

The process of understanding words can be conceptually divided into two parts: recognising the word and then understanding its meaning. Most scholars believe we have some sort of mental lexicon, where all the words are stored, both their forms and their meanings. The process of understanding words is then a process of accessing this mental lexicon. Two kinds of information are used: the perceptual/acoustic information, and knowledge of the context.

One of the first problems is to determine exactly what the word is – the incoming signal does not indicate words by putting gaps between them, as happens in writing. Naturally, there are a variety of acoustic clues, such as pitch, loudness and timing, but we also use more than the words themselves; we use our general knowledge to help us decide what the words are. Otherwise we could not explain the fact that we hear only those words that the speaker intends us to hear. If someone tells us about a visit from a guest, we never confuse this with guessed, because our knowledge of grammar tells us that one is a noun and the other a verb. A spoken sentence often contains many words that were not intended to be heard. That last sentence, for example, contains ‘Us poke can cent tense off in contains men knee words that were knot in ten did tube bee herd’ (Cole and Jakimik, 1980:139), but if we had heard the sentence, we would have recognised the right words from the topic of the discussion. Similarly, not only do we usually hear the right word, but we usually get the right meaning. For example, the word ‘printer’ has two meanings, and if we hear ‘the printer needs toner’, we know this is referring to a machine because of our background knowledge about how these printers work, and if we hear, ‘the printer is sick’, we know that this is referring to a person because people get sick and machines do not.

The effect of context seems to work through what has been called spreading activation. Words are connected by their features and their associations into a complex semantic network. So that ‘fire-engine’ will be associated with ‘fire’, ‘red’, ‘truck’, ‘fireman’ and so on. Each of these will in turn have their own associations, and in this way a complex network is developed. When a concept or word is encountered, the other things associated with it are also activated. Things closely associated will be strongly activated, and things weakly associated will be weakly activated. Once a word is activated it becomes easier to process that word.
There has been a lot of research into word recognition, and some of this is relevant to test construction: (i) higher-frequency words are recognised faster than lower-frequency words; (ii) word recognition is faster if the words are in a helpful context; and (iii) the fainter the sound, the longer it takes to recognise the word (Garnham, 1985).

Processing idea units

We do not usually recognise words in isolation, but as part of idea units. Processing idea units somehow requires taking the meaning of individual words, and combining these together to construct the meaning of complete utterances. Establishing the relationship between the meaning of individual words and the meaning of whole utterances is called parsing. Generally, the listener constructs a meaning of the idea unit, and then forgets the actual words and the syntax, and so only a summary of the meaning, the gist, remains (Sachs, 1967). Then the gist of each idea unit is combined with the gist of other idea units, and is then added to the gist of the text heard up to that point. In this way a summary of the whole text is built up.

Parsing idea units means determining the relationship between the parts of the utterance: i.e. who does what, to whom, and with what. This is based on both semantic and syntactic clues. The semantic information comes from such things as knowing that for a particular verb the subject must be animate, the direct object can be either animate or inanimate, and the instrument is usually inanimate. Syntactic clues can come from the word order, which in English is usually subject–verb–object; or from subject–verb agreement, which means that the noun which governs the verb is the agent.

The ease with which idea units can be understood depends on both syntactic and semantic factors. Some structures are generally more difficult to process than others: in English, negative statements take more time to process than affirmatives, and passive statements take more time to process than active statements. In semantic terms, plausible events are easier to process than implausible events. If the events are very plausible and very predictable, then simply paying attention to the content words may be enough to process the meaning. Making inferences based on the common relationships between content words is a very common comprehension strategy. But if events are not so plausible, or if it is necessary to disambiguate
the meaning, then the syntax is necessary to help determine the meaning. In the case of second-language listeners, syntactic skill tends to be low, and plausibility based on associations between content words is likely to play a greater role in language comprehension (Taylor and Taylor, 1990). Idea units are hardest to process when both the semantics and the syntax are challenging.

**Processing connected discourse**

Many people probably think of linguistic processing in terms of understanding sentences or short chunks of language. However, the rules and conventions of language use cover much larger linguistic units (for a review of research in discourse comprehension, see Graesser *et al.*, 1997). In typical conversations, speakers seldom organise complex thoughts into single utterances; instead, information may be spread out over many idea units and many conversational exchanges, as ideas are developed and expanded. Discourse level variables are important in the comprehension of spoken language.

**Cohesion**

One important variable is **cohesion**. Cohesion is a semantic relation between one element in a text and another element that has to be understood (Halliday and Hasan, 1976). One important type of cohesive device are **connectors** such as *however*, *therefore*, *despite that*, which indicate important semantic relations between the idea units they connect. In English, the definite article *the* is also very important – it indicates that the noun it modifies is a unique item already known to the listener and so, among other things, it serves to indicate what information is already shared between the speaker and the listener. Pronouns are another important type of cohesive device, and it is necessary for the listener to determine what a pronoun refers back to, in order to understand what it means. There are also pro-verbs – *do* in English or *suru* in Japanese, for example – that stand for other verbs, and are used to avoid repeating the main verb. The listener needs to understand which verb is being repeated.
Foregrounding

Related to the question of cohesion is the fact that not everything mentioned in a text is equally easy to refer back to; some things pass into the background while others remain in the foreground (Tyler and Marslen-Wilson, 1982; Anderson et al., 1983). Those things which remain foregrounded are active as a focus of attention, and are very easy to refer to; these are things that have been recently mentioned, or are related in some way to the content of the current utterance. Pronouns refer back to items which are foregrounded rather than those in the background. Sanford and Garrod (1981) suggested that the things that have been mentioned in the text are in explicit focus, and the things implied by that are in implicit focus. Thus, for example, if a house were mentioned, then that would be in explicit focus, and aspects of the house would be in implicit focus. So it would be possible to follow mention of 'the house' with a reference to 'the garden', and the listeners would be able to make sense of that and know which garden was being referred to.

Using world knowledge

It is already quite clear that any process of text comprehension presupposes a great deal of general non-linguistic knowledge about the world we live in, and how things work within it. This world knowledge can influence comprehension in several ways, but two are particularly important: (i) knowledge of the overall context restricts interpretations of the text; and (ii) knowledge of specific facts, or knowledge of how things usually happen, can be used to fill in details that are not explicitly stated in the text.

Inferencing

World knowledge is applied through the process of inferencing, and there is a wide variety of different types of inferences. For example, Hildyard and Olson (1978) classify inferences into three types: (i) propositional inferences are those that follow on logically and necessarily from any given statement; (ii) enabling inferences are related to the causal relationships between events or concepts; and
(iii) **pragmatic inferences** provide extra information which is not essential to the interpretation of the text, but which expands on it.

Inferences vary depending on how much background knowledge is needed to make the inferences, and these range from simple logical entailments, which follow on naturally from the meaning of words, to those that depend on very elaborate and extensive world knowledge. Inferences can also be classified according to how necessary they are to text interpretation: for example, **bridging inferences** must be made if a text is to be coherently interpreted, while **elaborative inferences** are not actually necessary to link together the parts of a text. Note also that inferences are not only made about situations described in the text, but can also be about the motives of the speaker, or the point the text is intended to illustrate.

Haviland and Clark (1974) and Clark and Haviland (1974) claim that the inferences necessary to understanding the meaning of a text, i.e. bridging inferences, are made while the text is being processed, although their research suggests that speed of comprehension is reduced when bridging inferences have to be made. Johnson *et al.* (1973) conducted a number of studies which suggested that elaborative inferences, i.e. those not necessary to understand the text but which add to it, are also made while a text is being processed, and are often incorporated into the comprehenders’ mental representation of the text and what it is about. They found that after a lapse of some time, comprehenders were unable to distinguish between what was actually in the text and the inferences they had made while processing it.

However, world knowledge is used not only to expand interpretation, but also to restrict it. When the general topic is familiar, knowledge about that topic can be used to interpret the text. The importance of this topic-specific knowledge has been established in a series of famous studies by Bransford and Johnson (1972, 1973). They gave readers a number of passages which were rated as incomprehensible by the readers; when a simple hint about the topic was given however, a title or a picture, the passages were rated as very comprehensible by the readers.

*Scripts and schema*

In order to explain how world knowledge is stored in memory and utilised in the comprehension of text, a number of similar and related
theories have been developed. Probably the most readily understood of these is the script theory of Schank and Abelson (1977). The idea is that an utterance sets up expectations, in the form of a script, about what normally happens in that context. The script is a mental structure which describes stylised, everyday situations; it is, in effect, ‘a very boring little story’ (Schank and Abelson, 1977:422).

Scripts are extremely numerous: we have restaurant scripts, birthday party scripts, football scripts, classroom scripts, library scripts and so on. In fact we probably have scripts for all the regularly occurring events in our lives. In listening to a story that calls up a script, the contents of the script automatically become part of the story, even if they are not spelled out. We know what happens in a restaurant, and there is no need to explain that the waiter brings the food, or that the customer pays the bill. Similarly, in a story about a Japanese public bath, the sento, we do not need to say that the bather undressed and went into a public room, washed thoroughly all over, and then got into a large communal bath of very hot water. This sequence of events is all part of the story because the sento script automatically includes it – that is what happens at the sento, everyone knows that, or at least everyone who has a sento script knows that.

This example raises a very important point, namely that scripts tend to be culture bound. We probably all have a bathroom script, but these may vary considerably from one culture to the next.

Related to scripts are schemata (Rumelhart and Ortony, 1977; Rumelhart, 1980); in fact scripts are really a type of complex schema. Schemata are structures for representing knowledge in memory, and are assumed to exist for most things we would want to represent in memory, including general concepts, situations, events, sequences of events, actions, sequences of actions etc. In some sense they may be thought of as representing stereotypes of these concepts, and in others they are somewhat like a play, with the internal structure of the schema corresponding to the script of the play. Like scripts, schemata guide the interpretation of text, setting up expectations for people, places or events. According to Rumelhart and Ortony, schemata are the basic building blocks of the human information-processing system, and are the key units in the process of comprehension.

It is clear that background knowledge is very important in listening comprehension. If the listener shares the same knowledge as the speaker, much of what is being said can be understood by means of