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;

iii 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.

Interpretations vary

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,