Promoting perception: lexical segmentation in L2 listening

John Field

This article calls for greater attention to the perceptual processes involved in second language listening—and particularly to the part they play in breakdowns of understanding. It suggests employing basic auditory phonetics as a means of classifying, diagnosing, and predicting problems of lexical segmentation. Recognition of how and why learners find speech input difficult to process can provide a programme of simple practice exercises which anticipate or rectify listening problems.

The importance of the signal

Two major considerations prompt this article. The first is a view that writers on second-language listening have perhaps concentrated overmuch on higher-level understanding. It is entirely appropriate that we should encourage second language learners to bring world knowledge to bear upon their listening experience; it is also important to provide adequate training in strategies which compensate for gaps in word recognition (Field 2000). But we should not lose sight of the primacy of the signal. We need to concern ourselves more than we do with speech as a physical phenomenon—with what English sounds like to the non-native listener and with the features which cause obstacles to understanding.

Some modicum of perceptual information, even if only a few words, is clearly needed before contextual knowledge can be brought to bear. But it is also a fact that many high-level breakdowns of communication originate in low-level errors. A second-language listener who hears I won’t go to London as I want to go to London, is making a small mistake based on phoneme discrimination. But this mistake may impact upon the interpretation of what comes next, and may even influence understanding of the text as a whole. Once learners have constructed a set of expectations for a text, they are notoriously reluctant to revise them, even if evidence comes in that contradicts them.

Inadequacies of the ‘comprehension’ approach

The second consideration is the dissatisfaction felt by many practitioners with the present approach to the teaching of second-language listening, based entirely upon the achievement of comprehension tasks. The comprehension approach is open to question for three reasons. Firstly it assumes that there is one ‘correct’ interpretation of an utterance, whereas comprehension theory (Brown 1995) tells us that there may be
more than one, depending upon the perspective of the listener. Secondly, it tends to home in upon small discrete points rather than checking global understanding. And, most importantly, it encourages us to focus upon the product of listening (in the form of answers to questions) but tells us nothing about the process.

It has often been said that we test listening rather than teaching it (Sheerin 1987). It is difficult to disagree with this assertion; but perhaps it misses the point. The fact is that the listening process is not accessible to inspection in the way that skills such as speaking and writing are. In order to establish if understanding has taken place, we have no choice but to employ what testing specialists term ‘indirect methods’: judging learners’ listening competence by their success or failure in answering questions or performing tasks. The problem, I suggest, lies not in the types of task we employ to check understanding, but in the use we make of the answers.

Suppose that Students A and B both give the correct answer in a comprehension exercise. We congratulate them and move on to the next question. What we have not taken on board is that Student A arrived at the answer as a result of understanding 96 words out of 100 in the text. Student B, on the other hand, understood only 20, but managed to achieve the correct answer by employing appropriate strategies. This is not to say that Student B is a worse listener than Student A. The point is simply that that these two students are listening in different ways, and need different types of support from the teacher.

It is essential that we follow up answers to comprehension tasks (both correct and incorrect ones) to find out how they were derived. Only in this way can we gain a clear picture of the strengths and weaknesses of our learners, enabling us to contribute constructively to their development as listeners. We should view the principal aim of a full-length listening session as diagnostic. It should provide us with insights into where understanding has broken down—insights which we can then follow up with small-scale remedial exercises which aim to prevent errors of interpretation (especially low-level errors) from occurring again.

A remedial approach to listening
A framework for analysing breakdown of understanding

A framework for classifying learners’ problems of understanding is provided by the levels of processing which feature in simplified psycholinguistic accounts of listening. These accounts view meaning as, in part, constructed through the ‘bottom-up’ assembly of larger and larger units. The signal is processed through several levels: auditory-phonetic, phonemic, syllabic, lexical, syntactic, semantic, propositional, pragmatic, interpretive. It is useful to regard these levels not only as steps in the construction of the message, but also as a series of stages at any one of which understanding can break down. It is often possible to determine that an error derives from inadequate processing at, for example, the phonemic, the lexical, or the syntactic level. To give examples: a learner might not be able to distinguish the difference between won’t and want (a phonemic problem), might not recognize the spoken form of the word vegetable (a lexical problem), or might overlook
the semantic implications of the phoneme /v/ in I’ve lived (a syntactic problem).

We need to be rather more persistent than at present in determining why breakdowns of understanding have occurred. Current practice encourages us to make easy but unfounded assumptions. Suppose an error occurs as the result of a learner failing to understand a particular word. The conclusion usually reached is that the word is not in the listener’s vocabulary. The teacher explains the meaning of the word, and feels that the problem has been resolved. Perhaps it has. But it could equally have arisen from several other causes, perceptual or semantic:

- Learner knows the word but attributed the wrong sense
- Learner failed to recognize a phonetic variation of a known word
- Learner knows the word in reading but not in spoken vocabulary
- Learner was unable to segment the word out of a piece of connected speech.

Exercise type

Once a problem has been identified, a set of simple exercises can be designed which raise awareness and provide focused practice. The most effective remedial listening work involves dictation. However, by ‘dictation’ is not meant the notorious dictée which many of us recall from our schooldays. Instead, the exercise should require that learners write down short, unpau sed sentences, each sharing the feature which caused the original difficulty; that these sentences are uttered as naturally as possible or taken from an authentic text; and that attention to errors of listening does not become side-tracked by attention to errors of spelling.

The remainder of this article discusses what is arguably the commonest perceptual cause of breakdown of understanding: namely, lexical segmentation, the identification of words in connected speech. The aim is to exemplify how low-level listening problems can be diagnosed by employing the basic knowledge of phonetics which most ELT practitioners possess. The difference is that the knowledge has to be stood on its head, so that we view phonetics not from the perspective of pronunciation practice but through the ears of the listener. Once identified, areas of difficulty can be tackled by means of simple 5-minute exercises; these might be remedial or they might anticipate problems of listening before they occur.

Lexical segmentation

We tend to overlook the fact that pauses in natural speech only occur every 12 syllables or so, which means that, unlike readers, listeners do not have regular indications of where words begin and end.¹ It is remarkable that we manage to separate out words within these 12-syllable chunks as consistently as we do.

Determining where word boundaries fall is a greater problem for the non-native listener than is generally recognized. A learner with limited English or weak listening skills adopts a strategy of scanning continuous speech for matches between sequences of sounds and items of known
vocabulary. In the anxiety to achieve matches, word boundaries are often breached:

Speaker: went to assist a passenger. Student 1 extracts sister

Speaker: the standard the hotel achieves. Student 2 extracts: stand at the hotel

We are all familiar with this phenomenon of grasping at cross-boundary straws and assuming words to be present which were not intended. (See Voss 1984 for examples.) The ‘matching’ strategy is a natural and productive one in the early stages of learning. The danger lies not in the strategy itself but in the tendency of students to overlook the tentative nature of the matches they achieve. There is a strong likelihood that Student 1 above will go on to construct a mental model of the text which includes somebody’s sister, even to the point of reshaping what comes next, in order to fit her in somehow.

So a first approach to lexical segmentation should demonstrate to learners the need for caution in word boundary allocation. One technique is to dictate ambiguous sequences, then to disambiguate them by adding additional words:

T dictates: an ice cream … [Ss write] … T continues dictation: a nice cream dress

T dictates: the boxes of … [Ss write] … T continues dictation: the boxes have been opened.

(For more examples, see Gimson 1994: 253.) This may seem a trivial exercise; but it is an effective way of demonstrating to learners that word boundary location may be a matter of guesswork, and that guesses may have to be revised in the light of later evidence.

Native segmentation strategies

A strategic approach to the lexical segmentation issue asks how it is that native listeners manage to locate word boundaries so successfully. Simple matching is not the answer. If it were, we would automatically begin a new word after, for example, hearing PORT in porter or PORTER in portable. The research of Anne Cutler and her associates (Cutler 1990) suggests that native listeners use a strong-syllable strategy, based on the premise that each stressed syllable marks the beginning of a new word. This strategy pays dividends. Using a corpus of spoken English, Cutler and Carter (1987) calculated that some 85.6% of all content words in running speech are either monosyllabic or stressed on the first syllable. The finding ties in with evidence from Hyman (1977) that lexical stress often fulfils a demarcative role. Many of the world’s languages have fixed lexical stress, which occurs on the first, the penultimate, or the final syllable of a word, and thus serves as a reliable cue to word boundaries.

From this, one might conclude that it is worthwhile to train learners of English to emulate the segmentation strategy adopted by native listeners. In fact, Cutler takes the view that it is impossible for learners to develop a segmentation routine in L2 which differs from the one used in their first language. However, her reservation refers to responses to the speech signal which are automatic. It does not rule out the possibility that
learners make slightly delayed decisions about what they hear, which resemble those of native listeners.

This, indeed, appears to be what many of them do. Learners show themselves sensitive to rhythmic regularities in the target language, and appear to learn from experience the value of inserting word boundaries before stressed syllables (Field 2001)—without being aware of what they have learnt. We should not be too surprised at this finding: infants acquiring English appear to use rhythm in the same serendipitous way to crack the code of connected speech (Jusczyk 1997).

With appropriate training, learners might acquire the technique much faster—though it may be necessary to train their ears to recognize lexical stress if it is marked differently in English from their native language. A challenging piece of authentic text might be played, and learners asked to write down stressed syllables and match them to words they know. Their attention can then be drawn to how many of these syllables initiate words. A similar awareness-raising exercise might involve playing recordings on low volume, and asking learners to transcribe the more salient syllables.

Focusing on stressed syllables in this way not only assists learners to locate boundaries but also draws attention to the fact that, in connected speech, such syllables are ‘islands of reliability’: louder and longer than unstressed ones. Some researchers (notably, Grosjean and Gee 1987) have even suggested that it is stressed syllables which serve to identify words for native listeners and that weak ones are accorded a different type of attention. On this analysis, our representation of the word *appear* is triggered by the sequence /pəər/, our representation of *indestructible* by /strʌk/. 

### Specific problems in the signal

If one major cause of segmentation problems is the lack of between-word pauses, a second and equally important one is the way in which the standard citation forms of words are modified when they occur in connected speech. Several different aspects will briefly be considered (reduction, assimilation, elision, resyllabification, and cliticization) and suggestions made for practice. For a detailed account of these phenomena, see Brown 1990.

### Reduced forms

Words, and even entire phrases, often appear in connected speech in a reduced form. One reason is that speakers economize on effort: for example, they avoid difficult consonant sequences by eliding sounds. Another reason is rhythmic: the patterns of English prosody dictate that certain closed class words such as prepositions, pronouns, and conjunctions are rarely stressed, and indeed that some may appear in a weak form (usually featuring schwa) in these unstressed contexts.

Unstressed syllables are shorter in duration, and less salient than stressed. They are also much less informative, because only two vowels, /ə/ and a shortened form of /ɪ/, predominate. Small wonder therefore that they pose perceptual problems for the foreign-language listener.
Three main types of reduction give rise to segmentation problems: contraction, weak forms, and the chunking of formulaic phrases. It is relatively easy to design a structured programme which introduces the second-language learner to these features as a listener. The proposal here is for 5-minute dictation sessions in which the sentences for transcription contain examples of particular types of reduced form.

A start can be made with the relatively simple area of contractions. Here, one goal is to get learners to recognize that the contracted verb is present at all. Assume that a class interprets *I've lived in London for 3 years* to imply that I no longer live there. It may well be that they have not understood the implications of this use of the Present Perfect. But the truth may lie at a much lower level: it may be that they have not noticed the presence of the */v*/.

Fifty-one function words in English possess alternative weak forms, most of them of high frequency. A second step in the kind of micro-listening programme proposed aims to ensure that the learner is able to recognize these words when they occur in connected speech. Very sensibly, many teachers choose to treat the weak form as the standard one, and the full form as the exception. This encourages the listener to construct a phonological representation which matches what is by far the more frequent form in connected speech. However, learners’ expectations of what they will hear are sometimes unduly influenced by exposure to the written language.

It is worthwhile covering the weak forms as comprehensively as possible (perhaps exemplifying four or five at a time). They can be presented again in read-aloud naturalistic sentences for transcription, where they should be given the kind of low prominence that they receive in spontaneous speech. For reference, a list of all 51 forms, mainly based on Gimson 1994, is provided in the Appendix.

When dealing with weak forms, it is important not to lose sight of the fact that we are presenting the material from the point-of-view of the listener, and not of the speaker. We also need to bear in mind that several of these forms are homophonous; the listener (whether native or non-native) needs to use syntactic context to distinguish them from each other.

<table>
<thead>
<tr>
<th>ə</th>
<th>a</th>
<th>əv</th>
<th>of</th>
<th>ən</th>
<th>an</th>
<th>əz</th>
<th>as</th>
<th>ə</th>
<th>yor</th>
</tr>
</thead>
<tbody>
<tr>
<td>ə</td>
<td>are</td>
<td>of</td>
<td>have</td>
<td>and</td>
<td>has</td>
<td>as</td>
<td>yo</td>
<td>your</td>
<td></td>
</tr>
</tbody>
</table>

Finally, we should recognize that native speakers often produce high-frequency sequences of words as chunks (Pawley and Syder 1983). These sequences may become very reduced, with phonemes and even whole syllables elided. They are only recognizable as a unit—and, indeed, it seems likely that native listeners store them as a single semantic and phonetic entity. It is good practice to dictate the most common of these formulaic phrases to learners, so that they can process them holistically.

*John Field*
when they encounter them. Favourites from my own collection are: /mɔ/ /naːpʃmɪtʃ/5

Penny Ur (1984: 46) provides a useful set of reduced sequences which can be used as material for exercises, as does Gimson (1994: 261–2).

It is not just the lack of pauses that makes it difficult to identify words in connected speech. Accommodatory phonological processes affect precisely the points at which the listener needs unambiguous information—namely word beginnings and endings. The most familiar of these processes are assimilation and elision. We tend to think of these phenomena as random, or at least as very complex. However, as Table 2 shows, assimilation is restricted in its operation, and quite systematic.

It is worth noting that assimilation in English is usually anticipatory, adjusting the ends of words in expectation of the sound that follows. The message for the learner is: trust the beginnings of English words rather than the ends. The sounds which are most subject to assimilation and elision are final /t/, /d/, and /s/. These, of course, provide many of the inflectional endings in English. Hence the irony of the grammar teacher telling learners to listen out for such endings, when they may be absent in spontaneous speech.

How to deal with the assimilation problem? Again by using dictation. The nine types of assimilation distinguished in Table 2 can provide the basis for a programme in which examples are either dictated as two-word sequences, or embedded in simple sentences.

| /n/ | [m] before [p, b, m] | ten people ➔ tem people |
|     | [ŋ] before [k, g]    | ten cars ➔ teng cars |
| /t/ | [p] or a glottal stop before [p, b, m] | that boy ➔ thap boy |
|     | [k] or a glottal stop before [k, g]  | that girl ➔ thak girl |
| /d/ | [b] or a glottal stop before [p, b, m] | good play ➔ goob play |
|     | [g] or a glottal stop before [k, g]  | good cause ➔ goog cause |
| /s/ | [ʃ] or omitted before [ʃ]  | this shirt ➔ thi shirt |
| /z/ | [ʒ] or omitted before [ʒ]  | those shoes ➔ tho shoes |
| /t, d, s, z/ | [ʃ, ʒ, f, j] before [ʃ] | Right you are ➔ rye chew are |

Did you go? ➔ di due go

Elision, unfortunately, follows a less consistent pattern than assimilation; but frequent examples such as didn’t ➔ [dɪnt] should certainly be practised for recognition in a connected-speech context or pointed out when they occur in a listening passage. We also need to pay special heed to the way complex clusters of consonants are elided:

next spring ➔ [nek’sprɪŋ]

Awareness of this kind of feature can aid learners in producing these clusters, as well as recognizing what has been omitted.

Lexical segmentation in L2 listening
Resyllabification and cliticization

Over the years, our ears become habituated to the vagaries of English. We rarely notice the effects which rhythm imposes upon words—effects which, for the learner, can considerably heighten the difficulty of recognition. Firstly, there is the process of resyllabification, where, in certain circumstances, a syllable-final consonant attaches itself to the following syllable:

\[
\begin{align*}
\text{went in} & \rightarrow \text{when tin} \\
\text{made out} & \rightarrow \text{may doubt (can’t) help it} \rightarrow \text{tell pit}
\end{align*}
\]

What complicates the situation for the listener is that, after resyllabification, words sometimes acquire false boundary cues. Thus, in the ‘went in’ example, the /t/ may well be lightly aspirated, suggesting that it is word-initial. Similarly, in made out, the removal of the /d/ from the first word is likely to be accompanied by a lengthening of the diphthong, so that it sounds to all intents and purposes like the open syllable may.

Secondly, there is cliticization: an effect which results from the way in which natural English speech tends towards a regular stressed-unstressed pattern. The preference of English speakers for the basic SW (strong-weak) foot means that they often attach two words for no reason other than a rhythmic one. This can happen in defiance of syntactic structure (Example 1). It can even lead to prefixes getting dislodged and being produced as if they were suffixes (Example 2).

Example 1: S W S W \rightarrow S W + S W

\[
\text{\textit{\text{\text{'gou t\text{\text{-}pause\text{\text{-}bed}}} \text{\text{-}pause\text{\text{-}}} \text{\text{'gout\text{\text{-}bed}} \text{\text{-}pause}}}}}
\]

Example 2: S W S W \rightarrow S W + S W

\[
\text{\textit{\text{\text{\text{'get k\text{\text{-}sartd}}} \text{\text{-}getk}}} \text{\text{-}sartd}}}
\]

How to handle these effects? They do not really lend themselves, like other segmentation problems, to short 5-minute dictation slots. The best advice is simply to be aware that they exist—and, when you encounter them in a listening text, to play and replay the relevant section to see if learners can puzzle out for themselves the correct distribution of phonemes and/or syllables.

Indeed, that is the message for all the perceptual difficulties described here. The important thing is to be aware of them, and to be prepared to practise them intensively if there are signs that they are preventing learners from identifying familiar words because of the special conditions of connected speech. The value of a signal-based approach of the kind described is that it draws our attention to problems of both perception and comprehension that would otherwise pass unnoticed.

Revised version received June 2002

Notes

1 Though accounts in the phonology literature often refer to allophonic cues as markers of word onsets, such cues are often absent in medium-paced spontaneous speech. (Gimson 1994: 265).

2 This is somewhat of a simplification. The criterion used by Cutler is not stress but the presence of a full-quality vowel.

3 They were unable to adjust this for frequency, but frequency-adjusted figures from a larger written corpus were very similar.
The shortened form of /u/ is much rarer.
more and more; Do you know what I mean?

References


The author

John Field researches L1 and L2 listening. He also writes and lectures on psychological processes in language, and this year has published two books on psycholinguistics (both with Routledge). He has experience as a teacher trainer and materials writer in many parts of the world, including Belgium, China, Hong Kong, and Tanzania, and has written English programmes for both radio and television. He currently teaches for the University of Leeds and for King’s College London. He is joint Co-ordinator of IATEFL Research SIG.
Email: jcf1000@dircon.co.uk
Grouped by word class (after Gimson 1994)

<table>
<thead>
<tr>
<th>word</th>
<th>weak form</th>
<th>word</th>
<th>weak form</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>ə</td>
<td>an</td>
<td>ən</td>
</tr>
<tr>
<td>any</td>
<td>ən</td>
<td>am</td>
<td>əm</td>
</tr>
<tr>
<td>some</td>
<td>əm</td>
<td>at</td>
<td>ət</td>
</tr>
<tr>
<td>the</td>
<td>ət</td>
<td>been</td>
<td>ən</td>
</tr>
<tr>
<td>for</td>
<td>fə</td>
<td>have</td>
<td>əz</td>
</tr>
<tr>
<td>from</td>
<td>ən</td>
<td>has</td>
<td>əz</td>
</tr>
<tr>
<td>of</td>
<td>əv, ən</td>
<td>had</td>
<td>əd</td>
</tr>
<tr>
<td>to</td>
<td>ət</td>
<td>does</td>
<td>əz</td>
</tr>
<tr>
<td>and</td>
<td>ənd, ənd</td>
<td>be</td>
<td>ðə</td>
</tr>
<tr>
<td>but</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>as</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>than</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>that</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>who</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>there</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>he</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>her</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>him</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>his</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>l</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>me</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>we</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>she</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>them</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>us</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>you</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>your</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
<tr>
<td>our</td>
<td>ðət</td>
<td>ðət</td>
<td>ðət</td>
</tr>
</tbody>
</table>