Language Awareness and the Teaching of Listening and Speaking

Christine C. M. Goh

Introduction

The teaching of second language (L2) listening and speaking has traditionally been practice-focused. Students participate in a variety of activities such as communicative tasks and listening comprehension exercises to practise their listening and speaking skills. In many classrooms, the two skills are practised in an integrated manner, typically with speaking taking precedence over listening. The teaching and learning goal of practice activities is predominantly on product, that is to say, how well learners can understand spoken input and express themselves, and their achievements are observed in degrees of success of comprehension for listening and fluency and accuracy for speaking. There is, however, another aspect of listening and speaking (oracy) instruction that shifts teachers’ and learners’ attention from product to process. This is the development of learners’ awareness about factors that influence their performance and how to manage these factors. A process-oriented approach to L2 oracy instruction has received far less attention in the classroom.

In keeping with the theme of this book, this chapter discusses such an approach by considering the role of learner awareness in the teaching and learning of L2 listening and speaking. It takes reference from the definition of language awareness as “a mental attribute which develops through paying motivated attention to language in use, and which enables language learners to gradually gain insights into how languages work” and its application in the adoption of “a pedagogic approach that aims to help learners to gain such insights” (Bolitho et al., 2003: 251). It also takes the view that different components of language sub-systems such as phonology, grammar and vocabulary are closely connected in the construction of meaning and that language awareness is “inseparable from text awareness” (p. 253). The meaning of ‘text’ in this chapter is synonymous with ‘discourse’: stretches of spoken language uttered in context. Discourse awareness underscores the principle that the production and comprehension of spoken language is a social and cultural act. In addition, language awareness also includes an awareness of the sociolinguistic milieu in which L2 oracy is taught and learnt.
This discussion of learner awareness does not focus only on the forms and use of language but also learners’ insights into the nature and demands of L2 listening and speaking, as well as strategy use that facilitates performance in the face of these demands. In this regard, the concept of metacognition (Flavell, 1979) will provide the theoretical framework for describing such kinds of learning processes engaged by language learners individually and collaboratively. Often defined simply as thinking about thinking, metacognition is recognized for its critical role in the success of many forms of learning, not least language learning (Wenden, 1991, 1998). Flavell (1979) defines metacognition as “knowledge and cognition about cognitive phenomena” (p. 906) and includes thinking about cognitive processes that one engages in, as well as ways to make these processes more efficient. Metacognition also involves “active monitoring and consequent regulation and orchestration of these processes in relation to the cognitive objects or data on which they bear, usually in the service of some concrete goal or objective” (Flavell, 1976: 232). With the right learning input and environments, language learners’ metacognition about language learning is accessible to introspection and reflection as well as amenable to changes and modifications (Wenden, 1998).

Awareness about the Nature of Speech

Teaching of listening and speaking has traditionally relied on using audio or video recordings for comprehension practice or communicative activities to encourage L2 learners to talk. While these activities help learners to practise listening and speaking, they do not usually provide opportunities to focus on the features of spoken English, and may even limit the learners’ opportunities to develop their linguistic and metacognitive knowledge about spoken English and oral interaction, because teachers and learners tend to focus on the outcomes, i.e. correct answers in listening comprehension or information and ideas communicated in pair or group work.

Listening and speaking are predominantly transient activities. Words are uttered in real time and comprehension of these typically impermanent words has to occur “on-line”. Even in formal presentations, speakers who have planned what to say will have to speak in real time. When no planning time is available, the challenge is even greater. Unlike writers, speakers cannot draft and redraft what they want to say before presenting a polished version of the spoken text to their listeners. The fluidity and spontaneity of spoken interactions would make this impossible. Learners have to think on their feet as they respond to questions or remarks from their interlocutors. It is impossible to prepare comprehensively for a conversation because the very nature of oral interactions means all participants have to negotiate and co-construct the discourse – many jokes have been told about hilarious outcomes of people regurgitating memorized responses to questions they thought they were being asked. Similarly, in listening, listeners seldom have an opportunity to hear what is spoken a second or even a third time. While some may be able to ask the speaker to repeat, listeners usually get just one chance to understand what is said.

Speakers often have to produce language under the pressure of time and this creates a large number of demands on their cognitive capacities (Segalowitz, 2010). It often limits their cognitive processing abilities and causes speakers, both proficient and less proficient ones, to attend to only selected aspects of their speech. Language learners will naturally have greater difficulty attending to structural aspects of language and
meaning simultaneously, forcing them to focus on getting their meanings across with whatever linguistic resources they have recourse to.

Teachers can help learners overcome some of these challenges through awareness of how spoken English works at different levels. Importantly, they need to know features of spoken language that make it different from written language and to be able to recognize them during listening. They also need to know what natural speech ‘looks like’ among competent speakers, so as to anticipate challenges in listening and set realistic expectations for their own performance. Teachers also need to bear in mind differences among various kinds of spoken texts that serve different social purposes and audience. In this way, they can help learners acquire a more thorough understanding of how spoken English works.

This section presents macro and micro features of natural spoken language, moving from general characteristics of speech and the patternings of spoken texts produced in particular social contexts to micro features of grammar, vocabulary and pronunciation (Goh and Burns, 2012). We begin with a comparison of spoken and written language as spoken and written texts fulfil different social purposes. Thus, even though they may be dealing with the same topic or content, spoken and written texts are constructed differently, drawing on different linguistic resources, namely grammar and vocabulary, to fulfill their functions according to the context. Nevertheless, spoken and written language do overlap in a number of ways, for example, the same vocabulary items may occur in both modes, some basic grammatical features remain the same, and discourse may be organized in similar ways for many similar communicative events. For example, texts concerned with making a proposal may adopt a problem–solution structure. Speech nevertheless has some grammatical and lexical features that differ from writing and the further a speech genre is from typical written genres, the greater these differences will be.

Comparing Speech and Writing

One way to examine the interrelationship between speech and writing is to consider language production as a continuum (Burns, Joyce and Gollin, 1996) where ‘pure’ spoken language and written language occupy each end. ‘Pure’ spoken texts are texts produced during immediate action in time and space (e.g. two friends talking about something they are looking at). These texts are interactional and co-constructed, unplanned and negotiated, and consist of less formal language. ‘Pure’ written texts are abstract, reflective and use formal language (e.g. a philosophy essay). Distinctions between speech and writing are blurred by texts that lie somewhere around the middle of the continuum. A panel discussion on TV has features of formal language and contains specialized vocabulary while retaining the interactional nature of speech. News broadcasts are predominantly monologic and contain many syntactic features of written language compared with a discussion; informal email and short text messages, on the other hand, are composed in a more speech-like manner.

Speech and writing also differ in their lexical density, which is the ratio of content words to function words in a clause (Halliday, 1985). Content words carry information and consist of nouns, adjectives, lexical verbs, and most adverbs. Function words show the grammatical relationship among content words and are needed to express meaning precisely. They consist of articles, prepositions, conjunctions as well as auxiliary and modal verbs. Compared with written language, spoken language has a lower lexical
density, that is to say there are more function words and the information in each clause is less packed. In contrast, writing shows greater complexity in terms of clausal embeddings and a higher degree of abstraction. Utterances for expressing the same meaning in speech also tend to have more words, as the examples adapted from Goh and Doyle (2014: 107) show:

**Spoken** (Total – 21 words: 11 content words (underlined) and 10 function words)

The book draws on a large number of studies to examine the ways in which children acquire and develop a language.

**Written** (Total – 17 words: 11 content words (underlined) and 6 function words)

Drawing on a substantial research base, the book examines the process of child language acquisition and development.

The spoken language example above is situated at a more formal point in the speech-writing continuum. While its lexical density is lower than the written version with the same meaning, it still has a semblance of a written sentence. Lexical density is lowest in informal conversations (Thornbury and Slade, 2006). Situated at the far end of the speech-writing continuum, informal or casual conversations epitomize the 'messiness' of speech because clauses tend to be added on one after another with the help of function words. Imagine a person explaining to her friend what the book is about. In such a casual interactional context, we may expect to hear a sequence of utterances such as these:

So this book I was reading // it’s about how children learn a language // how they acquire it // and how their language continues to develop // and to back up what he says // the author refers to a lot of research // but it’s easy to understand

Compared to formal speech and especially writing, casual speech rarely consists of well-formed and complete sentences. Utterances are produced in real time and linked to one another through the use of conjunctions (for example, so, and, but). The way speakers weave utterances together demonstrates a form of language complexity that is unique to speech, which Halliday (1985) referred to as grammatical intricacy. This adding on of clauses one after another is a common device that enables speakers to think as they speak and in some situations, it is a strategy for maintaining their turns. Large-scale corpora research such as the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) project on English speakers’ informal spoken data has shown this to be the norm (McCarthy, 1998). This has implications for teaching speaking, as many students and even teachers may expect all competent speakers to produce utterances similar to complete sentences in writing (Goh, 2009). With greater awareness about speech and writing, language learners will also learn to expect redundancies and repetitions when listening to a conversation or a lecture, or when interacting with other English speakers. Speech also displays less lexical variety, because the same words are often repeated in an interaction. Nevertheless, relexicalization or the use of synonyms of important content words is also a common feature of speech (McCarty, 1998). It is a way of affirming the topic of interaction and offering coherence to the text. On an interpersonal level of language use, repetition and relexicalization help to establish solidarity and agreement among participants in an interaction.
Features of Spoken Discourse

Spoken discourse, particularly when it is produced in day-to-day social interactions, has unique features as a result of common principles and demands of communication (Biber, Johansson, Leech, Conrad and Finegan, 1999). Students who are busy trying to understand what is said or focusing on what to say may overlook these features without the teacher’s help and miss out on linguistic knowledge that can help them become more effective listeners and speakers. It is important that teachers show them how language use is patterned and organized in different social contexts. Knowledge of these patterns or discourse routines (Bygate, 1998) will help learners predict what they hear and plan what to say.

In all interactions, someone who initiates talk with a question or a comment would expect an answer or a comment from the person they are talking to. These sequences are called *adjacency pairs* and consist of pairings such as question-answer, request-accept/decline, etc. If, however, the answer or comment is not what the questioner may have preferred to hear, the answerer will typically include a justification or an explanation. Findings from large-scale analyses of spoken corpora show that these sequences do not end with just the response but that the person initiating the sequence typically offers a comment on or an evaluation of the answer (Carter, 1998). Dubbed three-part exchanges, the third part is often filled by fixed or routinized phrases, such as ‘Really?’ and ‘That’s interesting’. The absence of a follow-up comment, Carter suggests, can make the interaction “cold and impersonal” (p. 44). Teachers should raise learners’ awareness of three-part exchanges to help them communicate naturally and appropriately. They can also be taught useful expressions to close the exchange loop. Their listening comprehension can also be facilitated if they can anticipate these moves and listen out for their interlocutors’ attitude or views to their responses. Here are two examples of three-part exchanges that can be used to illustrate this feature of interactional speech in class.

<table>
<thead>
<tr>
<th>1)</th>
<th>2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Fancy going to the movies tomorrow?</td>
<td>A: What did you do when you were in Malaysia?</td>
</tr>
<tr>
<td>B: Oh, I don’t think I can. I’ve got tons of work.</td>
<td>B: I went diving with some friends.</td>
</tr>
<tr>
<td>A: Oh, no worries.</td>
<td>A: That sounds like fun!</td>
</tr>
</tbody>
</table>

Other features of interactional discourse include talk management moves, such as introducing, maintaining and changing a topic, and turn-taking. These talk moves are facilitated by speakers’ discourse strategies. Two important strategies that learners need to learn are signaling openings/closings and offering feedback through fixed or routinized phrases and backchannelling. By teaching learners some appropriate formulaic expressions, we can help them be more confident in their spoken interactions. Backchannelling is the use of single words or short phrases (“Yeah, that’s right”) and sounds (“Uh huh, Uh-uh”) that are often accompanied by non-linguistic cues such as a smile, a nod or shakes of the head. Backchannelling enables listeners to indicate that they are interested in what is being said and they understand the speaker, as well as showing agreement, approval and other stances. The use of receipt tokens in the form of words and sounds demonstrates engaged listening, ensures smooth and cohesive interactions (McCarthy, 2003) and enables learners to maintain
an interaction, thereby staying longer and getting more opportunities for language input and practice (Gardner, 1998).

Genres of Speech

The concept of genre is found in different areas of language work such as English for Specific Purposes (ESP), systemic functional linguistics, rhetoric and language teaching (Paltridge, 2001). Genres are communicative events constructed for a specific social purpose. The term also refers to “the patterned discourse structures which have developed in a culture over time” and comprise transactional and interactional types (Burns et al., 1996: 2). Participants of a genre recognize its communicative purpose and establish the boundaries for what is permissible in terms of how the text is produced, what it addresses and how it is done (Paltridge, 2013). If interaction is successful, participants in these “episodes of speech” will go away satisfied with the engagement because the social encounter is what they have expected it to be (Carter and McCarthy, 1997: 8). On the other hand, if one or some participants interact and use language contrary to the nature of that particular genre, the other participants will feel frustrated, disappointed or even annoyed.

Genres of speech occur in diverse fields of language use and while some may overlap, others are distinctive to certain domains. ESP genres, for example, include “lectures, the graduate seminar, plenary lectures and poster session discussions at academic conferences, and college laboratory sessions” (Paltridge, 2001: 73), with lectures receiving the most attention in L2 listening (Flowerdew, 1994). Knowledge about the schematic structure of lectures can facilitate comprehension (Tauroza and Allison, 1994; Young, 1994), and ESP teachers can further help students to identify discourse patterns of lectures for specific disciplines because these patterns vary across disciplines (Dudley-Evans, 1994). Genres in social interactions include recounts of recent events, personal anecdotes and casual conversations (Burns et al., 1996), as well as specific types such as identifying or talking about oneself, comment-elaboration, debate and argument, service encounters, language-in-action, decision making, and learning and interaction in formal academic contexts (Carter and McCarthy, 1997). A lengthy sequence of talk of a specific genre can include one or more genre types framed by this larger generic structure. For example, in a dominant genre of extended casual conversations, we often find personal anecdotes, jokes, recounts or procedures produced by an individual speaker. These embedded genres or ‘chunks’ are identifiable genres with predictable discourse structures and are amenable to teaching (Eggin and Slade, 1997). Chunks can also occur in storytelling, sermons, social speeches, debates, show-and-tell and business presentations.

Knowledge of specific genres allows learners to activate their prior knowledge of scripts about similar spoken language events before and while listening (Vandergrift and Goh, 2012). This helps them anticipate each stage of an interaction as the discourse unfolds in predictable ways till it reaches its conclusion or achieves its final communication goal. Learners will also be able to free their attention from how to organize their speech so that they can select relevant content and convey their message through language that is appropriate and accurate. Freeing up cognitive space can be achieved through giving learners planning time, which research indicates can contribute to better overall speaking performance (Geng and Ferguson, 2013; Seifoori and Vahidi, 2012; Skehan and Foster, 1997, 2005).
Spoken Grammar

Research in corpus linguistics provided important insights into micro features of speech (see Biber et al., 1999; Carter and McCarthy, 2015; McCarthy and Carter, 1995) known collectively as spoken grammar. Evidence from large amounts of spoken data shows that micro features such as ellipsis, deixis, tags, tails, heads, and discourse markers are closely linked to speech production. Specific grammatical and vocabulary items are also found to be closely associated with different types of genres. Although there are areas of grammar in which written and spoken language overlap (for example, the past tense used in narratives and recounts when referring to actions), there are also distinctive grammatical features found only in specific speech genres. One such feature is the adding on of utterances or clauses by using conjunctions such as ‘and’ and ‘but’ highlighted in the discussion of grammatical intricacy previously. It is also common to see incomplete or fragmented structures (at least when compared to well-formed complete written sentences) added on to one another as the speaker attempts to fulfil a number of functions, such as elaborate, qualify and explain and keep the turn.

A major grammatical feature of speech is ellipsis, which is the omission of words and phrases because these vocabulary items are redundant in highly contextualized discourse. Listeners can infer what these are with knowledge established from earlier parts of the discourse or background knowledge on a topic. Such kinds of situational ellipsis are observed in casual talk where spoken English is informal, and less so in narratives because the narrator needs to be explicit and cannot assume too much common knowledge (Carter and McCarthy, 1997). We see this typically in short responses to questions:

A: How was the play last night?
B: Fantastic!

Another grammatical feature of speech is deixis, which comprises the “orientational features of language” pointing “backwards and forwards in a text as well as outside the text to a wider extra-textual context” (p. 13). Deixis is most common in conversations where interactants share common knowledge about the elements (for example, objects or persons) being referred to. Examples of highly contextualized language use include events where interactants are engaged in concrete actions together (cooking, playing, watching a game, etc.) or language-in-action (Carter and McCarthy, 1997) where meaning is clear from the shared knowledge of space, people and time. An example of language-in-action would be: “Turn this thing off and move it into that corner”. Non-linguistic cues, such as a tilt of the head or a gaze in a specific direction, also facilitate further contextual understanding for the listeners. Demonstratives (‘this’, ‘that’) and adverbs (‘here’, ‘there’) are used to respond to the spatial context while personal pronouns (‘she’, ‘it’, ‘I’) and adverbs of time (‘now’, ‘later’) point to the personal and temporal nature of talk, respectively.

The CANCODE data, mentioned earlier, revealed some unique grammatical features that vary quite considerably from canonical word order in written texts (Carter and McCarthy, 1997). Heads, tails and tags are almost exclusively associated with informal spoken English. Heads are nouns or noun phrases placed at the front of a clause to foreground their importance (“Your brother, he isn’t ever going to change, is he?”). Tails, which are words inserted at the end of a clause, are also used for emphasis and amplification (“Joan’s really kind she is”). Tags consist of an auxiliary verb and a noun or pronoun quite literally tagged on to the end...
of a declarative statement (“It’s hot in here, isn’t it?”). Typically a positive statement will have a negative tag and vice versa. Although appearing to be a question, tags do not always require a response. Rather, it is the speaker’s way of expressing a point of view to establish some solidarity with the listeners.

**Lexical Items in Speech**

The use of the word ‘thing’ in the example of deixis above illustrates the feature of vague language in speech, which is generally absent from writing that is formal and abstract. Instead of referring to the object as the ‘television’ or ‘TV’, the speaker uses a generic noun as it is clear from the context what the object is. Calling the object by name and the use of more precise language may make a speaker appear formal, pedantic or directive (Carter, 1998). Vague language is also often used to soften one’s speech: “Yes, what you did kind of irritated me” and to allow listeners some room to make certain decisions: “Could you give me something to drink?” Another lexico-grammatical feature of speech is found in the use of fixed or formulaic expressions. These expressions are used repeatedly in familiar contexts and typically at the beginning or ending of a stretch of discourse. They consist mainly of phrases and are pre-fabricated lexical chunks which serve many functions such as expressing an attitude or a stance, initiating talk, concluding an interaction, avoiding being too specific. Listeners can interpret the speaker’s general intentions based on these linguistic signals. Importantly, for speakers, formulaic expressions allow them to buy time for thinking and to maintain the interaction in a smooth and socially acceptable manner. Some well-worn expressions include “How’s it going”, “As a matter of fact”, “To my mind”, “You’ll never believe what/ who…”, “I’ll let you go”, “With all due respect”, “No offence intended”, as well as complete clauses, such as “Good morning and how are you today” and “You must be joking”.

Lexical items also serve the function of marking boundaries between topics and intentions and signalling the way discourse is organized. **Discourse markers**, as they are called, are a kind of “conversational punctuation” (Carter and McCarthy, 1997: 9), signalling speakers’ intentions such as a wish to initiate or close a topic, indicate reactions or move on to some other business. Corpus data has shown the significance of words such as “Right”, “Anyway”, “Fine”, “OK then” in serving these social purposes. Discourse markers that consist of whole phrases and formulaic expressions (“To move on to”, “To conclude”) signal stages in the macro structure of a text, indicating a change in topic or giving a point of emphasis as well as indicating intentions and logical connections at the clausal or micro level (Chaudron and Richards, 1986; Flowerdew and Tauroza, 1995). L2 listening comprehension is helped by knowledge of discourse markers (Jung, 2003; Eslami and Eslami-Rasekh, 2007) but they are only useful when learners can recognize these words in the streams of speech within the text. An important aspect of interpersonal relationships in language use is established through modality. This is another area where writing and speech overlaps. In speech, modality is established not only by the use of modal verbs (‘must’, ‘should’, ‘ought to’, etc.) but also modal adverbs showing possibility, obligation and other conditions (‘probably’, ‘likely’, ‘certainly’, etc.).

The significance of teaching certain grammatical and lexical features unique to speech has been challenged (Cook, 1998). Nevertheless, there are strong arguments against course books that use only scripted texts written in complete sentences, as they do not provide learners with the opportunities to encounter real texts outside the classroom, and learners need to have a sense of the kind of authentic ‘messy’ speech they are
likely to encounter in real-world communication and be prepared for it (Burns et al., 1996; Carter, 1998). A large survey of teachers and learners from various countries showed that while the learners were keen to conform to native speaker norms in speech, teachers in general did not think it was necessary (Timmis, 2002). Nevertheless, two-thirds of the teachers thought that awareness of features of spontaneous native speaker speech was important for their learners. This was echoed by teachers from Singapore and China (Goh, 2009), who showed a broad consensus of opinion that spoken grammar knowledge was useful for raising awareness about spoken and written language.

**Discourse Intonation**

Spoken discourse is constructed not only through lexical and grammatical choices but intonation, which is discourse sensitive, also plays a big part in shaping it (McCarthy, 1998). Speakers make intonational choices (stress and pitch movements) based on their assessment of the common (‘given’) knowledge they share with their speech partners and what information needs to be made explicit (‘new’) to them (Brazil, 1985/1997). Although this model of intonation is based on Southern British speakers and may not be completely relevant to speakers of English as an international language, teachers could still help learners become aware of how speakers including themselves can use intonation to convey and organize meanings for greater comprehensibility (Brazil, Coulthard and Johns, 1980). Phonological awareness activities can be based on the following features from Brazil’s discourse intonation (DI) model:

- **Tone unit:** Perceptible phonological blocks or segments in continuous speech which contain one or two syllables that are more emphatic than the others. These are called prominent syllables.

- **Prominence:** The highlighting of a word or words in a tone unit to indicate meaning selection from a range of possible choices. In bi- or poly-syllabic words, prominence is found in the stressed syllable. (e.g. TAble, comPUter).

- **Tone:** A major pitch movement within a tone unit. There are five tone choices: fall, fall-rise, rise, rise-fall, level. The two most frequent tones are fall (proclaiming tone) for new information and fall-rise (referring tone) for given information.

- **Key and Termination:** Pitch contrasts between prominent and surrounding words. There are three choices: high, mid and low.

Many authors have argued that pronunciation instruction should prioritize the teaching of suprasegmentals (stress, rhythm and intonation) because of their contribution to comprehensibility (e.g. Celce-Murcia, Brinton and Goodwin, 1996; Chun, 1988; Clennell, 1997; Morley, 1991; Pennington and Richards, 1986). Research shows links between speaker intelligibility and suprasegmental features (Tyler, Jeffries and Davies, 1988; Jenkins, 2002), speakers’ difficulties with primary stress or prominence (Hahn, 2004; Pennington and Ellis, 2000), and the overuse of particular tones (Pickering, 2001). Sensitivity to phonological features can also assist students in noticing important information during listening. When processing of aural input is limited by their working memory, L2 listeners are known to focus mainly on content words rather than function words because the former are more readily identifiable by prominence or primary stress in the utterances (Field, 2008a). English speakers in non-Anglophone countries may use prominence and tones differently from British speakers for similar pragmatic purposes (Goh, 2005), and this could have implications for initial crosscultural communication.
As part of learners’ language awareness, teachers can ask them to compare their intonation patterns with those of the other speakers of English.

**Metacognitive Awareness**

Metacognitive awareness for listening and speaking arises from learners’ introspection of their learning, as well as conscious attention to the nature and demands of listening and speaking processes and the strategies for facilitating these processes. Learners who are metacognitively aware have developed insights into how spoken language works during listening and speaking, and use this knowledge to help themselves become better at oral interaction and comprehension. In a process-oriented approach, teachers design learning activities that help students increase their awareness and learn how to use listening and communication strategies. A key component of metacognition is *metacognitive knowledge*. Flavell (1979) distinguishes three kinds of metacognitive knowledge – person, task and strategy:

- **Person knowledge** – How factors (e.g. age, aptitude, gender, and learning style) may influence learning; personal beliefs about oneself as a learner; personal learning problems and challenges.
- **Task knowledge** – The purpose, demands and nature of learning tasks; procedures and requirements for accomplishing these tasks.
- **Strategy knowledge** – Strategies that are likely to be useful for achieving specific learning goals; when some strategies may not be useful; how strategies can be used appropriately or in an orchestrated manner.

Another key component of metacognition is *strategy use*. Flavell (1976) highlights the importance of strategies for monitoring cognitive processes. Two other important strategies that manage cognitive processes are planning and evaluation. Together, planning, monitoring and evaluation are referred to as metacognitive strategies. Besides metacognitive strategies, we also need problem-solving strategies that work directly on language input to achieve learning goals or compensate for gaps in mental processing. These are often referred to as cognitive strategies. Affective strategies are needed for managing negative emotions while social strategies are useful for involving participants in an interaction to facilitate communication or getting more knowledgeable others to make a language task achievable.

Both metacognitive knowledge and the use of strategies are critical to L2 oracy development because learners do not normally get to ‘see’ the processes involved in their listening comprehension and speech production. During listening and speaking, their attention is mainly on extracting information and interpreting messages, as well as expressing their ideas clearly. Learners who developed their metacognitive awareness about listening through classroom instruction were more confident and self-directed (Goh and Taib, 2006; Graham, Santos and Vanderplank, 2011; Zeng, 2012). Learning to use strategies appropriately can also lead to better performance in listening (Vandergrift and Tafaghodtari, 2010; Zeng, 2014) and speaking (Jamshidnejad, 2011; Nakatani, 2005). Learners who develop their metacognitive knowledge in the classroom will also be more likely to use strategies in real-world communication (Zhang and Goh, 2006).

Teachers can use questionnaires or checklists to facilitate learners introspecting on their listening and speaking processes. The Metacognitive Awareness Listening Questionnaire
(MALQ) supports L2 learners’ introspection of their strategy use and themselves as L2 listeners (Vandergrift, Goh, Mareschal and Tafaghodtari, 2006). In ESP listening, the Academic Listening Self-rating Questionnaire (ALSA) can help raise learners’ awareness about their listening in an ESP context through self-appraisal of academic listening skills (Aryadoust, Goh and Lee, 2012). Inventories of skills, such as the proficiency level descriptors of the Common European Framework of Reference (CEFR) for speaking, have also been found to be useful to assist learners in introspecting on their speaking development and giving them the language to talk about it (Glover, 2011).

L2 speaking performance can also be enhanced by learners’ ability to use communication strategies (CSs) which are an important resource for addressing the social purposes and cognitive demands of oral interactions (Nakatani and Goh, 2007). CSs help learners fill communication gaps due to language problems and obtain help from their interlocutors, thereby enhancing their overall communication effectiveness (Faerch and Kasper, 1983; Nakatani, 2005). Learners can gain more time to produce speech, maintain their turns in an interaction and ensure that their conversations flow smoothly (Jamshidnejad, 2011; Maleki, 2007).

The pressure of real-time communication often means that learners do not have sufficient attentional resources to think about their own thinking. They may have a fleeting sense of what they are learning or doing and the difficulties they experience – something that Flavell (1979) termed as ‘metacognitive experience’ – but rarely do they have the luxury of time during an interaction to introspect on these experiences and examine them more closely. Even in classroom practice activities, learners usually need to produce communication outcomes under the pressure of time (for example, getting correct answers in a listening activity or speaking spontaneously in communication tasks), and there is little time and space for learners to learn their own thinking and communicating processes, or those of their classmates. Teachers therefore need to create the time and mental space for learners by including metacognitive awareness activities into practice-based listening and speaking lessons.

**Metacognition and Instruction**

Learners need to engage in activities where they can gain insights into how spoken language is used. These include language factors in spoken texts which we have earlier discussed, as well as mental and affective factors that they experience during listening and speaking. In this regard, Flavell’s metacognitive knowledge framework may help teachers to provide a reference for planning activities that can be integrated with listening and speaking practice materials. The following are instructional focus points associated with each knowledge type:

| **Person Knowledge** | • Personal factors that influence listening/speaking  
| | • Linguistic factors that affect listening/speaking  
| | • Listening/speaking problems, causes and solutions  |
| **Task Knowledge** | • Macro- and micro-features of speech  
| | • Top-down and bottom-up processes in listening  
| | • Planning, formulating and articulating speech  |
| **Strategy Knowledge** | • Listening strategies  
| | • Oral communication strategies  |
Learners should be aware of two key cognitive processes that occur interactively during listening comprehension, and the factors that influence these processes. The first is top-down processing, which is the application of prior knowledge or schema to infer and interpret meanings from a spoken text. They can learn to use a number of strategies, such as planning, focusing attention, monitoring, evaluating, inferencing, elaborating, contextualizing and predicting (Vandergrift and Goh, 2012: 277–284). Teachers can ask learners to prepare for a listening task by identifying strategies they can use according to the genres and context they will be engaging with. They can also work individually or collaboratively on metacognitive activities that promote self-regulation and self-appraisal (Goh, 2010).

The second cognitive process is bottom-up processing, which involves segmentation of streams of language into meaningful units so that message interpretation can occur. Also termed lexical segmentation (Field, 2008b), it is one of the main contributors to comprehension problems (Goh, 2000). Bottom-up processing is influenced by learners’ perception of segmentals such as phonemes of sounds, and this is largely facilitated by their ability to be attuned to the prosody or suprasegmental features of the L2. To become accustomed to the prosody of the language they are learning, learners need extensive exposure to the language. They may be able to get this from extended practice with spoken input in and beyond the language class (Renandya and Farrell, 2011). Nevertheless, it also calls for dedicated learning activities through which learners’ metacognitive knowledge of the demands of lexical segmentation can be examined and supported in class during listening lessons. For example, teachers can expand the last stage of the conventional pre-listening – listening – post-listening lesson structure to focus on phonological features of spoken texts and to review words and phrases that learners have difficulty recognizing during listening (Vandergrift and Goh, 2012). A suggested procedure for developing sound perception ability in the post-listening stage is as follows (Vandergrift and Goh, 2012: 158):

- Select a segment of the listening text and identify language features for focus
- Learners listen to it and transcribe it
- Show learners the original transcript and learners listen again
- Discuss how the features have contributed to listening difficulties
- Listen again without support

See Field (2008b) for activities for developing lexical segmentation.

Speaking involves a number of processes that occur interactively or sequentially and learners need to learn about these processes to deconstruct their own performance and manage each process as well as they can. Planning of the content or conceptualization forms the basis of speech production for without content there could be no speech. Information and ideas are encoded in the L2 using selected grammar and vocabulary in a process known as formulation. Finally, messages that have been encoded are expressed auditorily through articulation. These stages are based on the model of speech processing by Levelt (1989) which Bygate (1998) applied to L2 speaking. Apart from working to produce speech, teachers need to ensure that learners can monitor their speech and apply knowledge about discourse routines. Like listening, speaking in real time does not offer learners much time and space to monitor their speech or evaluate how well they
are doing. There is therefore a need for teachers to provide more explicit instruction on speaking that goes beyond communicative activities.

Burns (1998) proposed a discourse- or text-based approach to teaching speaking. It draws attention to the products of speaking in the forms of genres that learners will need to participate in and manage in authentic communication. It involves learners in a Teaching-Learning Cycle with four flexible phases carefully supported by the teacher: building the context, modelling and deconstructing the text, jointly reconstructing the text and independently reconstructing the text (Goh and Burns, 2012: 187; see also Burns et al., 1996). In each stage, teachers provide learners with opportunities to learn about the linguistic features of selected genres of speaking and produce language that is appropriate to the social purposes of each genre. They also teach students to use grammar and vocabulary that are relevant to each purpose and context. Another model is the Teaching-Speaking Cycle, which consists of a number of interrelated stages that enable students to combine practice and process-learning (Goh and Burns, 2012). The Cycle guides the teacher through stages for planning and focusing on language and discourse where learners pay motivated attention to the language features associated with a particular speaking task and appropriate communication strategies. Other stages in the Cycle encourage learner reflection and feedback from teachers. These stages are built around a main fluency-practice task which is repeated after the language awareness activity. By going systematically through these stages, learners can experience the process of learning to speak in a comprehensive and holistic manner.

Conclusion

Owing to the transient nature of spoken language, listening and speaking can be challenging for language learners. In many real-world communication events, learners may have just one chance to get it right. Oracy lessons should therefore provide learners with opportunities to not just practise listening and speaking but also understand the cognitive processes involved and how these are affected by the various features of spoken language.

Regular classroom practice may help learners feel more confident about speaking, but the effect of practising listening through exercises and activities is not always positive. As a matter of fact, one of the skills that beginning and intermediate language learners feel they are slowest at developing is listening, despite the amounts of practice they have. It is therefore critical that teachers teach learners how to listen and speak. A process-oriented approach develops learner awareness about how spoken language works in listening and speaking and at the same time promotes their oracy awareness of a conceptual and strategic nature. Such an approach points to the role of metacognitive knowledge that enables learners to understand the linguistic nature and demands of L2 listening and speaking processes and also gain deeper insights into strategies that can facilitate spoken language comprehension and production.

We now have substantial insights into features of spoken language as a result of extensive corpus research, while the debate of whether or not some of these insights are relevant to language learners is still an on-going one. Although a number of authors have argued in favour of including these insights into teaching materials, there is still a paucity of research into the teaching and learning of such features to substantiate the arguments (Timmis, 2012). More specifically, there is a need to examine the effect
of language awareness of these language features on listening and speaking, individually and collectively. In comparison, we have seen a noticeable increase in research into metacognitive awareness about listening in recent years, and some encouraging results have emerged. The same, however, cannot be said about speaking because there are few empirical studies that examine metacognitive awareness of L2 speakers. This area of research into the impact of language awareness on L2 speaking is therefore still a promising and exciting one.

Related Topics
Second language acquisition; assessment for learning; spoken discourse

References


