THE USE OF LISTENER RESPONSES IN MANDARIN CHINESE AND AUSTRALIAN ENGLISH CONVERSATIONS

Deng Xudong

Abstract

In recent cross-cultural studies of pragmatics, we have witnessed a rise in interest in the comparative study of phenomena beyond the level of single and decontextualised utterances encompassing those on the level of speech events such as casual conversations. The underlying premise for such studies is that different cultural groups may have different rules for participation in and interpretation of conversation that conflicts related to these rules are a major source of cross-cultural miscommunication. This study examines the use of listener responses by Chinese speakers in Chinese Mandarin conversations and by Australians in Australian English conversations. Following prior framework by Clancy et al. (1996), the study examines similarities and differences in the use of listener responses by these two groups of people in terms of frequency of use, types of listener responses, and the positions of listener responses with respect to transition relevance place. Results show that Australian and Chinese speakers do exhibit quite different conversational styles as evidenced in their use of listener responses. Specifically, while Australians use more listener responses, use a higher percentage of lexical expressions as their listener responses and tend to place their listener responses at a possible completion point, Chinese speakers use fewer listener responses, favour the use of paralinguistic vocalic forms as their listener responses and tend to place their listener responses during a turn. These findings may suggest a culture specific way of turn taking and of what it means to be polite in conversational behaviour.

Keywords: Listener responses; Mandarin Chinese; Australian English; Conversational styles.

1. Introduction

The conversational phenomenon of listener response has attracted a great deal of attention during the past four decades from such diverse scholarly disciplines as linguistics, conversation analysis, (cross-cultural) communication studies and experimental and social psychology. A number of terms have been used to describe this kind of listener behaviour, including ‘signals of continued attention’ (Fries 1952), ‘recognition’ (Rosenfeld 1966, 1967), ‘concurrent feedback’ (Krauss & Weinheimer 1966), ‘accompaniment signals’ (Kendon 1967), ‘listener responses’ (Dittmann & Llewellyn 1967, 1968; Bavelas, Coates, & Johnson 2002), ‘assent terms’ (Schegloff 1968; Leet-Pellegrini 1980), ‘back channels’ (Yngve 1970; Duncan 1972, 1973; Duncan & Niederehe 1974; Duncan & Fiske 1977, 1985), 'encourager' (Edelsky 1981), 'limited feedback' (Kraut, Lewis, & Swezey 1982), ‘responsive listener cues’ (Miller, Lechner, & Rugs 1985), 'minimal responses' (Fishman 1978; DeFrancisco 1991; Bennett & Jarvis 1991), ‘reactive tokens’ (Clancy et al. 1996), 'acknowledgment tokens' (Jefferson 1984, 1983/1993; Drummond & Hopper 1993a, 1993c), 'receipt tokens' (Heritage 1984) and
‘response tokens’ (Gardner 2001). In the present study, the term ‘listener response’ is used for the sake of its generality and easy comprehensibility.

The wide range of interests in the study of listener responses is not surprising as it is such a ubiquitous phenomenon in our daily conversations, but nonetheless one very seldom attended to by conversational participants in a conscious way. Thus the study of such conversational phenomena becomes even more important in cross-cultural communication settings. It is widely documented in cross-cultural communication literature that people from different cultures may follow a different conversation norm and can have different conversational styles, of which the use of listener responses in conversations is an important part. Interactional sociolinguists have shown repeatedly that when speakers with diverse conversational practices interact with each other, communicative difficulties or even miscommunication are most likely to occur, which can further result in negative cultural evaluations and stereotyping (Gumperz 1982; Tannen 1990).

This study compares the use of listener responses by Chinese and Australian speakers in their respective intracultural conversations. It aims to find out how they differ in terms of the frequency of the use of listener responses, types of listener responses used, and the location in which they place their listener responses in reference to a possible completion point of an utterance.

1.1. Approaches to the study of listener responses

In the history of research on listener response, two major strands of study can be identified which are representative of two different approaches to its study. One is the lumping approach, which treats as a single category or class a group of different forms of listener responses. The other is the splitting approach which is taken mainly by ethnomethodological conversation analysts. This approach analyses one or more discrete listener responses in their sequential context and tries to demonstrate that each token of listener response can perform distinctive interactional functions.

The lumping approach to the study of listener response is widely used in the fields of linguistics, language and gender, cross-cultural communication, and experimental and social psychologies. This approach probably starts with Fries’ (1952) study of American English sentence patterns, though earlier more general allusion to listener response could arguably have been made by Bales (1950) in one of his major groups of interaction process categories called ‘Positive Reactions’. Earlier studies of listener response in the lumping approach come largely from experimental and social psychology, commencing in the 1960s (e.g., Dittmann & Llewellyn 1967, 1968; Kendon 1967; Rosenfeld 1966, 1967) and continuing through the twenty-first century (e.g., Bavelas, Coates, & Johnson 2002). The earlier studies focused mainly on two general themes: The structural characterisation of listener response and its roles in conversation in general or more specifically in the conversational encoding and decoding process (see Duncan & Fiske 1985 for a different classification). A number of studies, which were concerned with the first theme, i.e., the structural description of listener response in the conversation, typically dealt with it in connection with its non-verbal versions such as headnods, gaze, and smiles (e.g., Birdwhistell 1962; Kendon 1967; Dittmann & Llewellyn 1967, 1968; Brunner 1979; Bavelas, Coates, & Johnson 2002). Others have mainly been concerned with its non-turn status in conversation (e.g., Yngve 1970; Duncan 1972, 1973; Duncan &

In addition to the structural description of listener responses, the study of their roles and functions in conversation, or in interpersonal communication more generally, is another frequent theme in the study of listener responses in the field of experimental and social psychology. This theme may have its origin in the study of the effects of feedback on human communication, a more general term which covers virtually all kinds of responses (visual or vocal) to a speaker ranging from headnods and smiling to interrupting and question-asking (e.g., Leavitt & Mueller 1951; Argyle, Lalljee, & Cook 1968; Rosenfeld 1966, 1967).

The other approach to the study of listener responses is what can be termed the splitting approach. This approach, mainly taken by conversation analysts, has examined some discrete listener response tokens in their sequential contexts. Unlike the lumping approach, it is not concerned with the relationship between the occurrences of listener response tokens and the external variables. Rather, it is mainly concerned with their occurrences with respect to the operation of the turn-taking organisation. In the conversation analytic literature, several listener response tokens have received an intensive systematic study. Each of them is found to be distinctive in terms of its placement and roles in the sequential environment and its consequences for subsequent turns. These tokens include: ‘yeah,’ ‘uh huh,’ and ‘mm hm’ (Schegloff 1982; Jefferson 1983/1993, 1984; Drummond & Hopper 1993a, 1993b, 1993c); ‘oh’ (Heritage 1984), assessments such as ‘wow’ and ‘good’ (Goodwin 1986), ‘okay’ (Beach 1993, 1995; Pillet-Shore, 2003), and ‘mm’ (Gardner 2001). Schegloff (1982), for example, studied listener response tokens like ‘uh huh’. He advises that discourse be studied as an interactional achievement, which is partially shaped by its turn-taking organisation. He observed that vocalisations like ‘uh huh’ in their sequential context can have two main and related usages: ‘a usage as continuer and a usage to pass an opportunity to initiate repair’ (p. 88). According to Schegloff, the most common usage of vocalisations such as ‘uh huh’ is as continuer, the function of which is to encourage the previous speaker to continue talking, and by producing ‘uh huh’, etc, the producer passes the opportunity to take a fuller turn at talk (p. 81). Listener response tokens like ‘uh huh’ occur frequently in an environment when an extended unit of talk by another is underway.

Jefferson (1983/1993, 1984) examined listener response forms like ‘mm hm’ and ‘yeah,’ which she labelled ‘acknowledgment tokens’. She found that the two tokens are functionally and sequentially different from each other in that the former (i.e., ‘mm hm’) indicates more of a passive recipiency and the latter (i.e., ‘yeah’) is more related to full speakership incipiency. Passive recipiency, according to Jefferson (1984), means that “its user is proposing that his co-participant is still in the midst of some course of talk, and shall go on talking” (p. 200). This is consistent with the above observation by Schegloff (1982) that listener response tokens like ‘uh huh’ and ‘mm hm’ serve mainly as continuers. Drummond & Hopper (1993a, 1993b) later took up the theme, attempting to reassess in a quantitative mode Jefferson’s claim about speakership incipiency of ‘yeah’ and passive recipiency of ‘mm hm’. Their studies received a critical response from Zimmerman (1993), although their findings reaffirmed Jefferson’s claim.

1.2. Classification of listener response

As studies of listener responses in the conversation analytic perspective seek mainly to
establish the uniqueness of individual tokens, the various classification systems of listener responses derive mostly from studies in the aggregate approach. A number of these systems have been based upon Duncan & his associates’ classification of ‘auditor back channel responses’ (e.g., Marche & Peterson 1993; for Duncan & his associates’ classification, see Duncan & Fiske 1985). Their classification distinguishes auditor back channel responses from other listening and speaking behaviours on the basis of the former’s non-turn status. That is, auditor back channel responses, according to Duncan & Fiske (1985), do not constitute a turn. The characterisation of listener response as a non-turn seems elusive since the concept of ‘turn’ itself has so far defied an adequate and clear definition. This has led Schegloff (1982) to the suggestion that “the turn-status of ‘uh huh’ etc. be assessed on a case-by-case basis, by reference to the local sequential environment, and by reference to the sequential and interactional issues which animate that environment” (p. 92: Note 16).

In dissatisfaction with the separation of listener responses from other speaking and listening cues purely on the basis of their status as a turn or non-turn, a few other researchers have sought to look into other criteria in the identification and classification of listener responses. These criteria include, most notably, the concept of ‘floor’ (e.g., Hayashi & Hayashi 1991) and the form and/or sequential organisation of listener responses (Tottie 1991; Clancy et al. 1996). While the use of the concept ‘floor’ for the classification of listener responses may suffer similar drawbacks as the use of ‘turn’, Clancy et al.’s classification system classifies listener responses (or Reactive Tokens in their terminology) ‘based partly on their form and partly on their sequential function’ (p. 354) and thus provides a promising link between the aggregate and the discrete studies of the conversational phenomena, making it plausible in comparing the cross-cultural use of listener responses while not ignoring the details of the sequential context in which they occur. The present study will adapt this system of classification in the comparative study of listener responses in Mandarin Chinese and Australian English conversations.

1.3. Cross-cultural study of listener response

Though scholarly interest in the conversational phenomenon of listener response dates back to the fifties, cross-cultural studies of this phenomenon are more recent, originating probably from the comparative studies of the conversational conventions of Japanese and Anglo-American native speakers of English (e.g., Lebra 1976; Hinds 1978; Clancy 1982; Mizutani 1982; LoCastro 1987; Maynard 1986, 1987; 1989, 1990; 1997; White 1989; Yamada 1989; White 1997; Hayashi 1988, 1990, 1991; Hayashi & Hayashi 1991; Clancy et al. 1996; Ward & Tsukahara 2000). The comparative study of the use of listener responses by cultural groups other than Japanese does not seem to receive much attention. A few exceptions include Lehtonen & Sajavaara’s (1985) study of Finnish conversation behaviours, Wieland’s (1991) study of the use of feedback between French native speakers and American advanced learners of French, Tottie’s (1991) study of British and American use of backchannels, Beach & Lindstrom’s (1992) study of Swedish and (American and British) English use of acknowledgment tokens, and Stubbe’s (1998) study of Maori and Pakeha’s use of verbal feedback in New Zealand English. These studies, by comparing the pattern of use of listener responses in different languages, have largely shown that speakers of different cultural groups use listener responses differently, which can result in communicative difficulties and negative cultural evaluations and
Listener responses in Mandarin Chinese and Australian English conversations

stereotyping.

1.4. Chinese and Australian use of listener response

Conversation behaviours of Australian speakers have very seldom been documented in the literature, let alone the use of listener responses in conversation (but see Thwaite 1993). This may be due to the underlying assumption that conversational patterns of Australian speakers resemble those of their American and British counterparts both in terms of the turn-taking patterns in general and in terms of the use of listener responses in particular.

As to the Chinese speakers, existing studies seem to show that they are very infrequent users of listener responses compared with some other cultural groups. Tao & Thompson (1991) compared the backchannel use of Chinese and Americans and found that Americans produce backchannels much more frequently than Chinese speakers and that Americans produce backchannels both during and at the end of the other party’s speaking turn whereas Chinese speakers produce backchannels exclusively at the end of and rarely in overlap with the other’s speaking turn. They also found that Americans’ use of backchannels includes the ‘continuer’ (Schegloff 1982), whereas Chinese speakers never use backchannels as continuers but use them to indicate understanding, confirmation, and acknowledgment of agreement. Another finding was that most of the backchannels produced by Chinese speakers were preceded by a noticeably long pause (longer than 0.3 seconds). Moreover, Tao & Thompson (1991) showed that the native speakers of Chinese (Mandarin) for whom English has become their dominant language exhibited more English backchannelling behaviour in terms of their form, frequency, and functions than Chinese backchannelling behaviour when speaking Chinese. While language transfer has usually focused on interferences from the native language to a second language, their result suggested interference in the opposite direction, from the second language to the native language.

Clancy et al. (1996) compare the use of reactive tokens in three languages - English, Japanese, and Chinese. They demonstrated that the three languages differ in the use of reactive tokens in several ways. Specifically, Chinese speakers use reactive tokens less than half as frequently as English and Japanese speakers, and they tend to use reactive tokens which are lexically contentful more than Japanese speakers, but a little less than English speakers. Chinese also place about the same percentage of reactive tokens at points of grammatical completion as English speakers but a higher percentage than do Japanese speakers. Clancy et al. (1996) suggest that Chinese listener behaviour, like that of the Indians of the Warm Springs community (or maybe that of the silent Finn), “is part of a ‘non-coercive cultural orientation’ that places high value on personal autonomy and avoids putting oneself above others” (p. 382). They also propose that avoidance of backchannelling by the Chinese reflects an appropriate stance of non-interference toward the speaker and represents an interactional style which values respectful deference on the part of Chinese interactants (pp. 382-383).

The scarcity of the use of listener responses by Chinese speakers was also noted by Günthner (1993) in her study of the conversational behaviours between German native speakers and Chinese speakers of German. With respect to the use of minimal responses such as ‘mhm’ and ‘ja’, she found that Chinese speakers so rarely used minimal responses in the course of the conversations that their German interlocutors often had to keep on
recycling part of their utterances, providing explanations, offering examples and initiating repairs, thinking that Chinese speakers had difficulty in understanding them (p. 288). German speakers, on the other hand, offered backchannel responses frequently when Chinese interlocutors were speaking. Günthner (1993) also observed the rarity of minimal responses in intracultural conversations between Chinese speakers.

Although previous studies have unanimously agreed that Chinese speakers are rare users of listener responses in conversation, the general validity of their findings may be questioned as they all suffer a number of methodological shortcomings, including: 1) Small sample size. Tao & Thompson (1991) used only two conversations for each of the languages studied (i.e., Chinese and English) and each conversation lasted for about five minutes. Clancy et al. (1996) used more conversations (eight altogether), but all the conversations were rather short with an average of less than three minutes in length. 2) Lack of control of group size. In both studies by Tao & Thompson (1991) and Clancy et al. (1996), the group size varies between two and three parties. 3) Lack of control of conversation participants’ gender and age. This is particularly true of Clancy et al.’s study, which did not specify the age groups of the participants and used conversations with random gender groupings. Though Günthner’s (1993) study used a much larger database (25 conversations in German and 3 conversations in Chinese), it specified neither the group size nor the gender groupings of each conversation. Further, the level of German of the Chinese participants may also pose a problem as it ranges from intermediate to fairly advanced. Most important of all, Günthner (1993) did not provide any quantitative information as regards the use of listener responses by Chinese and German speakers.

In view of the various drawbacks of the previous studies, the question remains: Is it really the case that Chinese speakers rarely use listener responses in their conversations? If so, how do they compare with Australian speakers in terms of the use of listener responses?

2. Data

The data used for this study are from 30 dyadic conversations: 15 Australian ones in Australian English and 15 Chinese ones in Mandarin Chinese. Both the Australian and the Chinese conversations consist of 5 pairs of female-female dyads, 5 pairs of male-male dyads and 5 pairs of male-female dyads. Each conversation lasted for approximately 20 minutes, but only a 10-minute segment was used for data. The segment was selected randomly, starting from the second minute onwards. The 60 participants (30 Chinese and 30 Australians) are university undergraduate students. The mean age of Australian students is 21.2 and that of the Chinese is 20.8. The Australian students are all Caucasian Australians who have received their primary or secondary education in Australia, and the Chinese students are from P.R. China. The conversation was between friends and was conducted in a classroom setting. The participants were given two conversation topics of general interest, but it was emphasised to them that they should feel free to talk about anything they liked.

3. Analytic framework for quantitative comparison

The present study will base its analytic framework upon Clancy et al.’s (1996) analytic
model in its quantitative comparison of listener response use in Australian and Chinese conversations. Their model sets out a well-defined classification of listener response tokens and an operationalisation of Sacks, Schegloff, & Jefferson’s (1974) concept of ‘transition-relevance places’. The former is useful to determine the frequency of listener response use and the preference of some type(s) of listener response tokens over the other(s); and the latter helps specify in a more systematic and empirically viable way the location in which listener response tokens tend to be placed. But a number of modifications and elaborations will be made in line with the theoretical stance of this present study.

3.1. Types of listener responses

Clancy et al. (1996) distinguished between five types of listener response tokens (or “Reactive tokens” in their term): Backchannels, Reactive Expressions, Collaborative Finishes, Repetitions, and Resumptive Openers.

3.1.1. Backchannels (BC)

Backchannel (BC thereafter) is defined by Clancy et al. (1996) as “a non-lexical vocalic form, [which] serves as a ‘continuer’ (Schegloff 1982), display of interest, or claim of understanding” (p. 359). While I would emphasise BC’s non-lexicality in form and its passive recipiency in sequential function, I would like to treat the latter half of their definition with reserve as the characterisation of BC as display of interest and claim of understanding appears too general and equivocal. As Schegloff (1982: 79) rightly argues, any utterance produced by one speaker following that by another exhibits an orientation to, or an attention to, it. Thus according to him, to say that a listener response token displays attention to preceding talk does not help discriminate it from any other talk, or tell us what a particular token does or can do, and therefore why a participant might choose to produce it rather than something else (Schegloff 1982: 79).

With the definition of BC having been clarified, some exclusions can be quite straightforwardly made. First, any lexical items such as ‘yeh’ and ‘right’ in Australian English and ‘dui’ (‘right’) and ‘shi ah’ (‘yes + vocalic particle’) in Mandarin Chinese are not to be taken as BC. Second, not all non-lexical forms are necessarily BCs. Only those which are used by the interlocutors to show their recipiency are taken to be BCs. Thus for example, if the vocalisation serves as the second pair part of an adjacency pair (Sacks, Schegloff, & Jefferson 1974), e.g., as an answer to a question or as a response to an offer, it is not considered a BC as it constitutes a full turn by itself. This also applies to other types of listener response tokens. Further, non-lexical vocalisations which serve as assessments such as ‘wow’ in Australian English and ‘ai ya’ and ‘wa’ in Mandarin Chinese (Goodwin 1986; Goodwin & Goodwin 1987, 1992a, b) are not coded as BCs but will be coded as Reactive Expressions (cf Clancy et al. 1996).

Typical BC forms in each of the two languages found in the data are listed in Table 1 (but with their prosodic variations omitted from the table).
Table 1 shows that Australian and Chinese speakers use similar types of BC forms in their respective intracultural conversations. This may suggest that the two languages, i.e., Australian English and Mandarin Chinese, share to a considerable extent similar interactional resources in the performance of a non-primary speaker’s work in a conversation. But it should be remembered that these listener response tokens, though similar in form in the two languages, may not be in strict conformity with each other in pronunciation and in interactional functions. For example, the ‘oh’ in Mandarin Chinese, which sounds like a shortened ‘or’ in Australian English, is at least not the same in pronunciation as its Australian counterpart ‘oh’.

3.1.2. Reactive Expressions (RE)

Reactive Expressions (RE thereafter) was defined by Clancy et al. (1996) as “a short non-floor-taking lexical phrase or word” that a non-primary speaker produces in response to the primary speaker’s talk (p. 359, my emphasis). For this definition, I would like to make two additions. First, assessment-type non-lexical forms (e.g., ‘wow’ and ‘oh wju’ in Australian English and ‘ai ya’ and ‘wa’ in Mandarin Chinese) will also be included in the category of RE. Second, those lexical phrases or words which preface a full turn (i.e., those lexical phrases or words used to display imminent speakership) are also taken to be REs. Thus the revised definition of RE for this study would be: A short free-standing or turn-incipient lexical phrase or word, or an assessment-type non-lexical form, produced by a recipient in reaction to the speaker’s talk. Typical REs in Australian and Chinese conversations in this study are listed in Table 2 (again the prosodic variations of these REs are omitted).

Table 2 Typical Reactive Expressions in Mandarin Chinese and Australian English

<table>
<thead>
<tr>
<th>Mandarin Chinese</th>
<th>Australian English</th>
</tr>
</thead>
<tbody>
<tr>
<td>shi (shi) [copula]</td>
<td>yeh (yeh yeh)</td>
</tr>
<tr>
<td>shi ah (shi ah) [copula + final particle]</td>
<td>yeh oh yeh</td>
</tr>
<tr>
<td>ah shi ah</td>
<td>yeh hm</td>
</tr>
</tbody>
</table>

1 The bracketed part indicates the permissible range of such tokens used in the actual conversational data. Thus, for example, hm (hm hm hm) under Mandarin Chinese means that up to four ‘hm’s have been used by Chinese speakers in their conversations.
**Listener responses in Mandarin Chinese and Australian English conversations**

<table>
<thead>
<tr>
<th>Mandarin Chinese</th>
<th>Australian English</th>
</tr>
</thead>
<tbody>
<tr>
<td>hhmhm shi ah</td>
<td>yeh (yeh) sure (yeh)</td>
</tr>
<tr>
<td>shi ah, jiu shi ah [indeed+final particle]</td>
<td>yeh yeh for sure</td>
</tr>
<tr>
<td>jiu shi</td>
<td>yeh right</td>
</tr>
<tr>
<td>jiu shi ah</td>
<td>yeh (yeh) exactly</td>
</tr>
<tr>
<td>ah jiu shi ah</td>
<td>yeh excellent</td>
</tr>
<tr>
<td>shi ma? [copula+question marker]</td>
<td>ah yeh</td>
</tr>
<tr>
<td>shi ba [really]</td>
<td>hm yeh</td>
</tr>
<tr>
<td>na shi [that+copula]</td>
<td>oh yeh (yeh)</td>
</tr>
<tr>
<td>na shi ah</td>
<td>yes</td>
</tr>
<tr>
<td>na dao shi [that’s true, though]</td>
<td>yeesh</td>
</tr>
<tr>
<td>ah na dao shi</td>
<td>(oh) right (yeh)</td>
</tr>
<tr>
<td>eh na dao shi</td>
<td>that’s right</td>
</tr>
<tr>
<td>zhe dao shi [this is true, though]</td>
<td>hm yeh that’s right</td>
</tr>
<tr>
<td>na ye shi [that’s also true]</td>
<td>yeh that’s right yeh</td>
</tr>
<tr>
<td>ye shi ah</td>
<td>ah of course (yeh)</td>
</tr>
<tr>
<td>dui (dui dui) [right]</td>
<td>oh excellent</td>
</tr>
<tr>
<td>ah dui (dui dui dui dui)</td>
<td>true</td>
</tr>
<tr>
<td>oh dui (dui dui dui dui dui dui dui dui)</td>
<td>shit</td>
</tr>
<tr>
<td>ah dui ah</td>
<td>hm shit</td>
</tr>
<tr>
<td>ah dui de</td>
<td>oh shit</td>
</tr>
<tr>
<td>dui ah</td>
<td>shit, that’s good</td>
</tr>
<tr>
<td>dui ya</td>
<td>ah bullshit</td>
</tr>
<tr>
<td>ai dui</td>
<td>(oh) really?</td>
</tr>
<tr>
<td>ai dui le</td>
<td>did they?</td>
</tr>
<tr>
<td>ah dui, shi ah shi ah</td>
<td>are you?</td>
</tr>
<tr>
<td>eh dui dui dui</td>
<td>is it?</td>
</tr>
<tr>
<td>dang ran le [sure]</td>
<td>have you?</td>
</tr>
<tr>
<td>na dang ran [that’s for sure]</td>
<td>oh have you?</td>
</tr>
<tr>
<td>ai ya, tian la [oh, my god]</td>
<td>he has?</td>
</tr>
<tr>
<td>ah na tai hao le [ah that’s excellent]</td>
<td>was she?</td>
</tr>
<tr>
<td>hao hao hao [good good good]</td>
<td>was it?</td>
</tr>
<tr>
<td>zhen de? [really]</td>
<td>you did?</td>
</tr>
<tr>
<td>wa</td>
<td>no (no no)</td>
</tr>
<tr>
<td>ai ya</td>
<td>oh no</td>
</tr>
<tr>
<td>ai yo</td>
<td>oh no yeh</td>
</tr>
<tr>
<td></td>
<td>hm no</td>
</tr>
<tr>
<td></td>
<td>na</td>
</tr>
<tr>
<td></td>
<td>hm na</td>
</tr>
<tr>
<td></td>
<td>no wa</td>
</tr>
</tbody>
</table>
The simple length of the list of REs found for the two languages in our data seems to show that Australian speakers produce a greater variety of REs than their Chinese counterparts in their respective intracultural conversations. This may suggest to a certain extent that Australians favour the use of listener responses in conversation more than Chinese speakers.

3.1.3. Collaborative Finishes (CF)

A Collaborative Finish (CF thereafter) was an utterance produced by the non-primary speaker to finish a previous speaker’s utterance (Clancy et al. 1996: 360). Collaborative sentence construction in conversation has been extensively researched by Lerner (1989, 1991, 1996). CFs are found in conversations of both Australian English and Mandarin Chinese in my data. I will give one example each for the conversations of the two languages (Ex 1 from the Chinese conversations followed by a largely literal English translation of the example and Ex 2 from the Australian ones)².

Ex 1: C5ff: aCF1

B: wo ben ren shen xihao wenxue ah. xingxiang siwei
A: hmm.

B: qiang yidian, [wo jiu xuyao wo nan pengyou shi yige, A: [hm.

² The transcription conversation employed by conversation analysts, originally developed by Gail Jefferson and thus called the Jeffersonian Transcription System, was adopted for this study with slight modifications and a horizontal layout rather than the original vertical layout. This system was employed because it attends to the minute details of a conversation, making it optimal to analyse various subtle features of listener responses such as placements of listener responses. For a detailed description of this transcription system, see Sacks, Schegloff, & Jefferson 1974: 731-734; Schenkein 1978; Psathas & Anderson 1990).
Listener responses in Mandarin Chinese and Australian English conversations

→ A: **lixing siwei qiang [yidian de]**.
   B: [dui, lixing siweil qiang yidian de].

Translation

   B: I myself very much like literature. and am better at
   A: hmm.

   B: thinking in images, [so I’d like my boyfriend to be someone,
   A: hm.

→ A: **who is better at logical [thinking]**.
   B: [right, better at logical thinking].

**Ex 2: A3ff: bCF1**

   A: and she’s one of those women. you don’t know whether

   A: she’s like-,

→ B: **nice or not**.

3.1.4. **Repetitions (RP)**

When the non-primary speaker repeats a portion of the speech of the primary speaker, it is coded as a Repetition (RP thereafter). Again I will give one example each for the conversations of the two languages (Ex 3 from the Chinese conversations and Ex 4 from the Australian ones).

**Ex 3: C7mf: mRP3**

   F: Lin Yutang:: you shenme shu ah hai you?
   M: Lin Yutang:::

   M: Zhongguoren.
   F: Zhongguoren mei kan. wo juede haoxiang

   F: you yi ben Lin Yutang de shenme:: sanwen ji ah.

→ M: .hhh **sanwen ji**.

Translation

   F: what other books did Lin Yutang:: write?
   M: Lin Yutang:::

   M: The Chinese.
   F: haven’t read The Chinese. it seems to me
F: that Lin Yutang has something like a collection of essays.

→ M: .hhh a collection of essays.

Ex 4: A5ff: bRP1

A: I finally I did read it, but um, it was absolutely

A: crappy, stupid thing↑.
B: who’s it by? (0.4)

→ B: Hardy. right.
A: Thomas Hardy.

3.1.5. Resumptive Openers (RO)

Resumptive Openers (RO thereafter), according to Clancy et al. (1996: 362), share the same form as BCs, i.e., they are both non-lexical vocalisations. But RO is distinguished from BC in that RO is followed by a full turn whereas BC is free-standing. Normally there is only a short pause between a RO and the full turn following it. In the sequential context, RO serves to “acknowledge the prior turn and commence a new turn” (Clancy et al. 1996: 364). Thus unlike BCs, which serve as ‘continuers’, ROs signal speakership incipiency.

3.2. Transition Relevance Places

The notion of ‘Transition Relevance Places’ (TRP) is discussed here as a reference point for specifying the location of listener response occurrences. Specifically, with reference to TRP, we can determine whether speakers tend to place their listener response tokens at either (1) points of possible transition from one speaker to another or (2) during another speaker’s turn.

The concept of TRP was first brought out by Sacks, Schegloff, & Jefferson (1974) in their study on the systematic organisation of turn-taking in conversation. According to them, turns are made up of turn-constructional units and the units are syntactically defined (i.e., sentences, clauses, phrases, words). The end of a turn-constructional unit is a TRP, where turn-change from one speaker to another normally occurs. Although Sacks, Schegloff, & Jefferson pointed out the syntactic nature of turn-constructional units, they did not spell out exactly how they are actually realised linguistically in the conversation context.

To overcome this indeterminacy of TRP, Clancy et al. (1996) and Ford & Thompson (1996) propose the concept of Grammatical Completion Point or Syntactical Completion Point (SCP), which is in practice equivalent to Sacks et al.’s TRP. They set out several criteria for the recognition of SCP (or TRP), which are summarised as follows:
1) An utterance is considered syntactically complete if, in its sequential context, it could be interpreted as a complete clause, i.e., with an overt or directly recoverable predicate, without considering intonation or interactional import.

2) Syntactically complete sentences can always be extended through further additions, so points of syntactic completion may be incremental.

3) Syntactically complete utterances include elliptical clauses, answers to questions, and backchannel responses.

I will illustrate this first with an example in the Chinese data:

**Ex 5: C5ff**

A: wo::wo guan cha/ guo/ le/.

I::I observe/ PRT/ PRT/.

I::I have already observed.

Here in this example, A’s utterance contains three syntactic completion points (indicated by slashes). The first SCP occurs after ‘cha’, which means that the clause before ‘cha’ is complete. The second SCP comes after ‘guo’, a final particle indicating an action already finished. It does not mean that the word ‘guo’ by itself constitutes an independent unit, but that the whole clause up to and including ‘guo’ is complete. The same procedure applies to the third SCP, which occurs after ‘le’, also a final particle.

The following example from the Australian data is an utterance containing a series of SCPs (again indicated by slashes).

**Ex 6: A1ff**

B: and I wrote it./ and er you have to like hand some of the stuff in/ so the teacher can have a look/ at it↑./ (. ) and she was sort of, reading/ and stuff./ and she criticised my whole poem./ (.0.2) except for one line/ she ticked,/ like gave me a double tick↑./

In addition to the notion of syntactic completion point, Ford & Thompson (1996) introduced the terms of pragmatic and intonational completion points to form what they call ‘Complex Transition Relevance Places’. But these two terms have proven not as useful as SCP. For intonational completion points, which are defined as ends of intonation units with a final contour, most of these points fall on SCP in Chinese conversations, thus providing little new information as regards the location of listener response occurrences. This may be because Mandarin Chinese is more of a tone language than of an intonational one, which makes it difficult for an untrained ear to distinguish between a final and a non-final intonation. The same is true of pragmatic completion points, which are a combination of intonation and conversational action sequencing. Therefore, unlike Clancy et al.’s study (1996), which used syntactic and intonational completion points, the present study will use only the syntactic completion point (or TRP in Sacks, Schegloff, & Jefferson’s terminology) in the specification of the location of various listener response tokens.
3.3. Additional specifications in the application of the framework

Before the framework can be applied to the data for this study, two further specifications need to be made. The first is to do with frequency counts and the second with the whole spectrum of TRP coverage.

We will first look at the issue of frequency counts when two or several listener response tokens occur in close proximity to each other in conversation. Are these tokens to be taken as separate instances or one single complex instance? In this respect, I will adopt the criterion used by Tottie (1991) in her corpus study of backchannel use in British and American English conversations. The criterion she uses is the relative proximity of these backchannel items to each other. If two or several backchannel items are adjacent in time, they would be regarded as one backchannel. On the other hand, if they are separated by several words or by a long pause, they would be taken as separate backchannels. To be more exact, I would propose that if two or more listener response tokens are separated by two or more words, they be regarded as separate cases. Otherwise, they would be taken as a single case. In the following three examples from the Australian data, the first two (i.e., Ex 7 & Ex 8) are examples where the two listener response tokens are coded as one single instance and the third (i.e., Ex 9) is one in which the two tokens are considered two separate instances.

**Ex 7: A12mm: aRE6**

B: I did half of-, like nearly all of my-, video, .hhh on Edit

B: Suite-, (0.3) the good one, Edit Suite [s- ] seven,

→ A: [>yeh<] >yeh.<

B: and then, for that last shot, where he’s jumping off the cliff↑, ...

**Ex 8: A13mm: bRE13**

A: there was not one book in this library on, what I needed for

A: education, and but actually, Curtin has a [journal ] article [but,

→ B: [yeh. ] [yeh.

A: .hh I could have gone and got from there.

**Ex 9: A13mm: aRE25, 26**

B: and it’s a big chance for all the Nazis to jump on the bandwagon,

B: (. ) ban practically anything, from TV. ex[cept,] what they deem

→ A: [ye:h.] [ye:h except=[and then you’re getting] into censorship again.
Now we come to the issue of TRP coverage, i.e., when should a listener response token be regarded as occurring at a TRP? As was mentioned before, Clancy et al. used what they called the Complex Transition Relevance Places (CTRP) as a reference point for the placement of listener response tokens. They counted a token as occurring at a CTRP “if it occurred in the clear (i.e., not in overlap) immediately after the CTRP” (Clancy et al. 1996: 365). But if this over-rigid concept of CTRP is to be applied to the TRP, it would necessarily fail to capture a whole range of spots which interlocutors systematically use for the placement of their listener response tokens. According to Jefferson (1983), TRP should not be taken as one fixed spot but it can fall along several different points near that of possible completion:

there is some flexibility as to what ‘at’ a possible turn-ending is, which is why we talk of a transition place instead of a transition point” (Jefferson 1983: 3, original underlining).

Thus I will use the same principle to distinguish between listener response tokens occurring at transitional points and those occurring at nontransitional points. Specifically, I make the length of two syllables a demarcation point between transitional listener response tokens and nontransitional ones. That is, if a listener response token occurs more than two syllables away from any transitional relevance places, whether it occurs in the clear or in overlap, it will be counted as occurring at a nontransitional point (cf. West & Zimmerman 1983; Clancy et al. 1996). Otherwise, it will be counted as occurring at a TRP. In the following example from the Australian data, the five listener response tokens (one BC and 4 REs) produced by B are all counted as occurring at a TRP.

**Ex 10: A1ff: bRE2, 3, 4, 5 & bBC2**

```
A:  O::H got this stupid thing at work. .hh[h it’s called super crew↑.
1 → B: [yeh.

B: heh [hh heh heh heh
A: [and you know they try to make us work ha:rd. .hh anyway

A: there’s incentive right ’cos they’re all different stations in my

A: [work↑=like how you cook all the bur[gers an’- .h[hh
2&3→ B: [yeh. [yeh. [what do

4 → B: you do? [yeh.
A: um I do front counter↑. .hh like ser[ving people in

A: dining room and fries and s[tuff like that. .hhhh like oːh you
5 → B: [hm.

A: will not believe it, er it’s so craːss
```

In contrast, the listener response tokens in the following examples (again taken from the
Australian data) are all counted as occurring at a non-TRP.

Ex 11: A11mm: bBC1

A: also I suppose um, that would be sh-that would really be sort
A: of u::[m, more the distance assassination. rather than the sort
→ B: ["hm"
A: of-, (0.2) the back of the head, close range kind of deal.

Ex 12: A2ff:: bRE14

B: but then again, hh would he be classed as a second year or
→ B: first year. [yeh. ↑first year. ]
A: fir-he shou-[he would’ve tried out with all fir]st years.

4. Results and discussion

4.1. Overall frequency of listener responses

Figure 1 shows the number of listener responses that each of the 60 Australian and Chinese speakers produced in a 10-minute conversation. It shows quite clearly that the Australian speakers (with a mean of 44.2) uttered many more listener responses than Chinese speakers (with a mean of 26.4). The Mann-Whitney \( U \)-test results confirm the significance of the difference between these two groups of speakers in terms of the total number of listener responses they produced in the conversation \( U (30, 30)=1; p<0.001; 2\text{-tailed} \).

This result adds strong evidence to the findings by many previous researchers that Chinese speakers are relatively rare users of listener responses in conversation - in comparison with speakers of many other cultures such as Americans (e.g., Tao & Thompson 1991; Clancy et al. 1996), Japanese (Clancy et al. 1996; Mizuno 1988 [cited in Clancy et al 1996]; Liu 1987 [cited in Clancy et al 1996]), Germans (Günthner 1993) and now Australians.

Another observation we can make from Figure 1 is that there is a large individual difference within each cultural group, as indicated by the Standard Deviation numbers (18.1 for the 30 Chinese speakers and 19.2 for their Australian counterparts). These individual differences can serve to mitigate any cultural generalizations we may want to make. In other words, not all Chinese speakers are reluctant users of listener responses while at the same time not all Australian speakers are ready to insert many listener responses in interaction with other conversationalists.
4.2. Distribution of listener response types

Table 3 compares the distribution of the five types of listener responses in Australian and Chinese conversations. The most striking difference between the two groups of speakers is in the relative frequencies of BCs and REs they use. For the Chinese speakers, the most favoured type of listener response is clearly BC, which occupies 44% of all the listener response tokens they uttered. RE is the second, comprising about a third of all the listener response tokens they used (29%). But for the Australian speakers, RE is obviously their favourite type of listener response with about two thirds of all their listener response tokens being RE (63%). BC, on the other hand, is the distant second, taking up a little more than a fifth of the total listener response tokens they produced (22%). The rankings for the three minority types of listener response are similar in the two languages with RO being the third and CF and RP being the remote fourth and fifth.
The finding that the Chinese speakers favor the use of BC over the more lexical types of listener responses agrees well with that of Clancy et al.’s study (1996), in which the Chinese speakers’ use of BC comprises 47.2% of all listener response tokens. But unlike the Americans in Clancy et al.’s study (1996), who also prefer the use of BC (37.9% of the total number of listener response tokens), the Australians in this study use the RE predominantly (63%) while only producing the vocalic BC about one fifth of all their listener responses.

This discrepancy between the result of this study and that of Clancy et al. (1996) with respect to the distribution of the different types of listener response tokens (and more specifically to the use of RE) can be due to both methodological and cultural factors. Methodologically, the two studies differ from each other mainly in that whereas this study adopts a definition of RE which includes not only free-standing tokens but also turn-incipient ones, RE in Clancy et al.’s study (1996) may have only included free-standing lexical items or expressions (pp. 359-360). This definitional difference for RE (and perhaps that for BC as well) may account for part of the discrepancy in this respect. Another likely reason may be cultural. Though both are native speaker of English, Australians and Americans can have distinctive conversational styles, as in the case of two ethnic groups in the US who also exhibit the differential use of conversational strategies (Tannen 1981a, 1981b, 1984).

### 4.3. Placement of listener responses

<table>
<thead>
<tr>
<th></th>
<th>Transitional</th>
<th>Nontransitional</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>596 (75%)</td>
<td>196 (25%)</td>
<td>792 (100%)</td>
</tr>
<tr>
<td>Australian</td>
<td>1153 (87%)</td>
<td>173 (13%)</td>
<td>1326 (100%)</td>
</tr>
</tbody>
</table>

Table 4 Number and percentage of transitional and nontransitional listener responses
This result does not seem to agree very well with that of Clancy et al.’s study (1996). In their study, Chinese speakers place a higher percentage of their listener responses at TRP (or Grammatical Completion Points in their terminology) (88%) than American English speakers (78%). What is most noteworthy here is that a higher percentage of transitional listener responses were found for the Chinese speakers in their study (i.e., 88%) than in mine (75%). This can be due to the different types of data we collected. The data they collected consisted of eight very short conversations with no control of gender groupings and group size as well as no specification of participants’ age and social status, whereas the data for this study are all extended dyadic conversations between friends of similar age and social status. But two other factors may also come into play which may have given rise to this mismatch. The first one is again the definitional differences for the term TRP between the two studies. The concept of Grammatical Completion Point in Clancy et al.’s study (1996) is much narrower than that of TRP in my study. In their study, they may have only counted a listener response as occurring at a TRP when it occurred in the clear (i.e., not in overlap) immediately after the TRP. In this study, however, we count a listener response as occurring at TRP whenever it falls within two syllables of a possible completion point, whether it occurs in the clear or in overlap. It is possible that this technical difference in terms of the definition of TRP may be part of the reason for the two different results obtained by the two studies. One other possible reason behind this mismatch can also be the smallness of the database in Clancy et al.’s study, which finds only 35 listener responses altogether for their eight Chinese conversations (as compared with the 792 listener responses for the 15 Chinese conversations in this study) (see Fig. 1 in Clancy et al.’s study).
5. Conclusion

I compared the use of listener responses by Australian and Chinese speakers in their respective intracultural conversations. The comparison was made in three aspects: Overall frequency of listener responses, preference of listener response types and the placements of listener responses with reference to a possible completion point. In all these three areas, differences have been located between the two groups of speakers. Firstly, Australians use significantly more listener responses than Chinese speakers; secondly, Australians prefer to use linguistic lexical expressions such as ‘yeh’ and ‘right’ as their reaction to the primary speaker’s ongoing talk whereas Chinese speakers favour the use of paralinguistic vocalic forms such as ‘hm’ and ‘ah’; and finally, while both Chinese and Australian speakers place a great majority of their listener responses at a possible completion point, Australians place a higher percentage of their listener responses at possible completion points than Chinese speakers, and Chinese speakers, on the other hand, place a larger proportion of their listener responses during a turn than do their Australian counterparts. These differences may strongly suggest a culture-specific pattern in the use of listener responses, which is an integral part of turn-taking mechanism and conversational styles. In what follows, I discuss turn management characteristics of Chinese and Australian speakers with particular reference to their use of listener responses and based on these descriptions, I draw implications with respect to cross-cultural communication.

The use of listener responses by Chinese speakers has not received a great deal of attention. The few available studies have all revealed that Chinese speakers are relatively rare users of listener responses in comparison with speakers of some other languages such as Japanese, Americans and Germans (e.g., Tao & Thompson 1991; Clancy et al. 1996; Mizuno 1988 [cited in Clancy et al 1996]; Liu 1987 [cited in Clancy et al 1996]; Günthner 1993). The findings of this study are consistent with these previous studies that Chinese speakers use listener responses infrequently, this time in comparison with Australian speakers. Even when a listener response is used, it tends to be a paralinguistic cues, a minimal vocalic sound, placing relatively little burden on the part of the primary speaker (Clancy et al., 1996). The tendency of the Chinese speakers to utter more listener responses during the turn of the primary speaker rather than at a possible completion point may be linked to the little use of eye contact with fellow conversationalists, as suggested by many anecdotal evidences. Studies on the functions of gaze direction in dyadic conversations have provided evidence that eye gaze is closely linked with the production of listener responses (Kendon 1967/1990; Bavelas, Coates & Johnson 2002).

The differential use of listener responses on the part of the Chinese speakers as opposed to their Australian counterparts may have implications in their interaction with other cultural groups of people such as Australians who may favour more frequent use of listener responses, the use of more lexically contentful reactions and the placement of these responses at the end of a turn. The differential conversational behaviour may well be subjected to different culture-specific interpretations. For Chinese speakers, their infrequent use of listener responses and their predominant use of vocalic sounds as reactions to the speaker can be interpreted as adopting a non-interference stance toward the speaker and showing respect or deference for the speaker to formulate and produce his or her talk undisturbed (Clancy et al. 1996) and may also indicate their
preferential use of negative politeness strategies that respect the speaker’s negative face (i.e., the need to be independent, have freedom of action, and not be imposed on by others), in opposition to the positive politeness strategies that attend the positive face (i.e., the need to be accepted and liked by others, treated as a member of the group, and to know one's wants are shared by others) (Brown & Levinson 1987; see also Lakoff 1973). For Australians and other likely frequent users of listener responses such as Germans, Japanese and Americans, failing to provide listener responses may be interpreted as lack of cooperation and involvement, which may have contributed, at least partially, to the often perceived Western images of the inscrutable Chinese (Young 1994) and of the passive and unresponsive learners of a foreign language ( Günthner 1993).

The use of listener responses by Australian speakers has also received very little attention, probably due to the assumption that Australian English speakers resemble American or British English speakers in their conversational styles, at least to the extent that they may follow similar turn-taking mechanisms as postulated by Sacks, Schegloff & Jefferson (1974). Nevertheless, there is some evidence for distinctively Australian ways of speaking. For example, Guy et al. (1986) noted that Australian English speakers often used a high-rising intonation in statements called Australian Questioning Intonation. With respect to the use of listener responses, Australian speakers, based on the findings of this study, would expect frequent insertions of listener responses, the majority of which would likely be lexically contentful and occur at a possible completion point. This conversational style may indicate Australian speakers’ preference of using positive politeness strategies and camaraderie (Lakoff 1973; Brown & Levinson 1987). But for Chinese speakers, the excessive use of listener responses on the part of Australian speakers may be taken as imposing, pressurizing, and disruptive.

While emphasising the strong culture specificity in the use of listener responses between Australians and Chinese, we are at the same time not denying many other factors which influence their use. For one thing, gender plays different roles in the use of listener responses in conversations of the two languages (see Deng 1999). For another, great variations have been observed between individual speakers, indicating the existence of within-culture individual differences in the use of listener responses. Furthermore, as the subjects of the study are all university students within the age range of 17 to 26, whether the results can be generalised to other population groups is not known. Finally, there may also be other contextual factors such as the choice of conversational topics and the degree of intimacy between the speakers which can also influence the use of listener responses.

As the use of listener responses constitutes only one aspect of a speaker's conversational style, other aspects such as pace and pausing, the use of overlaps, and even non-verbal cues such as eye gaze are also in need of study so that the whole picture of the speaker's conversational style can be revealed. Natural conversational data in intracultural communicative situations could be further compared with the results from this study. Finally, intercultural conversational data could be analysed to see whether differences in conversational styles in intracultural communication do lead to communication difficulties in intercultural encounters.
References


Drummond, K., & R. Hopper (1993c) Acknowledgment tokens in series. *Communication Reports* 6.1:
47-53.


Listener responses in Mandarin Chinese and Australian English conversations


Ward, N., & W. Tsukahara (2000) Prosodic features which cue back-channel responses in English and
328  Deng Xudong


