

MSU Working Papers in Second Language

Studies Volume 2, Number 1

2010

Running head: ACCENTEDNESS, COMPREHENSIBILITY, TONE

Accentedness, comprehensibility, and tones in the speech of second language learners of Chinese

Submitted for review to *MSU Online Working Papers in Second Language Studies*

on 3 May 2010.

Abstract

Research has shown that lexical tones pose considerable challenges for second language (L2) learners of Chinese (Wang, Jongman and Sereno, 2003; 2006) and accurate production in tones is crucial for intelligibility and comprehensibility of Chinese L2 speakers' utterances (Chen, 1997). This study investigates the interrelationships among degree of foreign accent, comprehensibility, and tone production, and attempts to identify the extent to which tone production contributes to degree of foreign accent. The participants were 30 adult L2 learners of Chinese (10 English, 12 Japanese, and eight Korean native speakers) who were enrolled in a Chinese as a Second Language Program at a large university in China. The participants responded to a semi-direct oral proficiency interview and the oral responses were used for evaluation of degree of foreign accent, perceived comprehensibility, and quality of tone production. ANOVA, correlation, and regression analysis were applied to data analysis. Results show that degree of foreign accent strongly correlated with comprehensibility, and comprehensibility moderately correlated with tone production. Tone production was found to be a poor predictor of perceived accentedness. Results suggest that segmental and other suprasegmental aspects of L2 speech might have outweighed the contribution of tone production to foreign accent.

Key Words: accentedness, comprehensibility, Chinese tones, L2 speech

Accentedness, comprehensibility, and tones in the speech of Chinese second language learners

Introduction

Second language (L2) learners vary substantially in the degree to which native listeners perceive their speech to be accented. Munro and Derwing (1995a) define *accentedness* as the degree to which the pronunciation of an utterance sounds different from the native speaker norms. Studies that examine factors contributing to degree of foreign accent suggest a wide variety of variables that influence the production of accented L2 speech, such as age of onset learning (AOL) and length of residence in an L2-speaking country (Flege and Fletcher, 1992; Flege, Munro, and MacKay, 1995; Flege, Yeni-Komshian, and Liu, 1999; Piske, MacKay, and Flege, 2001; Thompson, 1991). It is generally concluded that accentedness is an inevitable aspect of a late L2 learner's speech, even among learners who have been immersed in the L2 environment for many years or have achieved high proficiency in the L2 (e.g. Bongaerts, 1999; Flege, 1988; Riney and Flege, 1998).

The main objective of this study was to investigate a distinctive suprasegmental aspect of speech—Mandarin Chinese tones (tones, hereafter)—and its association with measures of accentedness and comprehensibility. The domain of tones was selected for investigation for three main reasons. First of all, acquiring accurate tone production is a broad acquisitional task involving not only the development of a new phonological system at the suprasegmental level, but also the memorization of each individual lexical tone, both of which require a heavy cognitive load on the part of L2 learners of Chinese (Brown-Schmidt and Canseco-Gonzalez, 2004; Chen, 1999; Wang and Kuhl, 2003). Secondly, accurate pronunciation in tones is crucial

for accurate information and for intelligibility and comprehensibility of Chinese L2 speakers' utterances (Chen, 1997; Sun, 1997). Thirdly, research in Chinese second language acquisition (SLA) suggests that tones pose considerable challenges for late L2 learners (Chen, 1999; Ye and Connine, 1999; White, 1981; Winke, 2007), particularly for those from non-tonal L1 backgrounds (Chen, 1997; Feng, 2004; Tao and Guo, 2008; Wang *et al.*, 2003; 2006). Therefore, tones provide an ideal case for the study of foreign-accented speech among L2 learners of Chinese with non-tonal L1 backgrounds.

Literature Review

Research that examines the degree of foreign accent often relates the properties of accentedness with *comprehensibility* and *intelligibility* (Derwing and Munro, 1997; Flege *et al.*, 1995; Kennedy and Trofimovich, 2008; Munro and Derwing, 1995a; 2001; Munro, Derwing, and Morton, 2006; Trofimovich and Baker, 2006). Comprehensibility is defined as a listener's estimation of difficulty in understanding an utterance while intelligibility refers to the extent to which a speaker's utterance is actually understood (Munro and Derwing, 1995a; Munro *et al.*, 2006). Munro and Derwing's studies consistently find that degree of accentedness is related to comprehensibility and intelligibility and suggest that foreign accent is a salient indicator of a nonnative linguistic background (Munro and Derwing, 1995a). Research also shows that the interrelationships among accentedness, comprehensibility, and intelligibility are complex and often confounded by various other factors, such as listener's familiarity with accents (Bent and Bradlow, 2003; Derwing and Munro, 1997; Gass and Varonis, 1984) and the speaker's or listener's first language (L1) background (Bent and Bradlow, 2003; Munro *et al.*, 2006).

The interwoven relationships among accentedness, comprehensibility, and intelligibility have been examined in a number of studies (Munro and Derwing, 1995a; 1995b; 2001; Munro *et*

al., 2006). Munro and Derwing (1995a) found that although the strength of foreign accent correlated with perceived comprehensibility and intelligibility, a strong foreign accent did not necessarily entail decreased comprehensibility or intelligibility. They claimed that even if an L2 speaker had a strong accent, his or her speech could still be highly intelligible. Munro and Derwing (1995b) further revealed that although a strong foreign accent did not necessarily interfere the speaker's intelligibility, native listeners needed extra effort or processing time to understand accented L2 speech, which led to a higher level of difficulty in comprehension. Derwing and Munro (1997) also found that listeners tended to award poor comprehensibility scores to accented speech than the listener's intelligibility score would warrant.

For late adult L2 learners, non-native patterns of pronunciation are pervasive, manifesting both at the segmental (Anderson-Hsieh, Johnson, and Koehler, 1992; Derwing, Munro, and Wiebe, 1998; Munro, 1993) and suprasegmental (Derwing *et al.*, 1998; Munro, 1995; Munro and Derwing, 1998; Trofimovich and Baker, 2006) levels of speech production. Several studies have investigated the relationships among different prosodic aspects of L2 speech (such as intonation, speech rate, pausing features), and perceived foreign accent and comprehensibility (Anderson-Hsieh *et al.*, 1992; Flege, 1988; Jilka, 2000; Munro, 1995; Trofimovich and Baker, 2006; van Els and de Bot, 1987). Munro and Derwing's study (2001) indicated that speech rate exhibited a strong relationship with listeners' judgments of accentedness and comprehensibility.

Trofimovich and Baker (2006) found that duration of pauses and speech rate were associated with the degree to which the L2 speech was perceived as accented. Jilka (2000) investigated the contribution of intonation to foreign accent among German native speakers learning English and English native speakers learning German. He found that intonation played a crucial role in foreign-accented speech. Other studies (e.g. Anderson-Hsieh *et al.*, 1992; Munro and Derwing,

1995a; Rasier and Hiligsmann, 2007) consistently reported that segmental errors have a less detrimental effect on listeners' judgments of comprehensibility and accentedness compared to suprasegmental errors—thereby underscoring the importance of suprasegmental characteristics of L2 speech.

Chinese characters are exclusively monosyllabic. Each Chinese character phonologically consists of segments (consonants, vowels) and the suprasegmentals, tones (Chao, 1948; Sun, 1997). Studies in acoustic characteristics of Mandarin tones often classify tones with two sets of descriptive terms: one denoting pitch height (high, low) and the other pitch contour (rising, falling) (Wang *et al.*, 2003). Mandarin Chinese phonemically distinguishes four tones that serve to provide contrasts in word meaning. These four tones can be described as high-level, high-rising, falling-rising, and high-falling, respectively, based on the tones' fundamental frequency variation patterns (Fu and Zeng, 2000; Howie, 1976; Wang and Sereno, 1999). Tones are manifested at the level of the syllable, the smallest structural unit for carrying prosodic features (Shen, 1990). The pitch contours of the four tones are often indicated with the tonal diacritics, such as \bar{a} , \acute{a} , \check{a} , \grave{a} . Many Chinese characters are minimal pairs with respect to tones. For example, “yàoshi” means “if” and “yào shǐ” means “key”, respectively. The difference in tones between the second character “shi” marks the difference in meaning. Thus accurate pronunciation in tones is essential to convey the precise meaning.

Tonal processing system and production mechanisms are unique to tonal languages (Hao and de Jong, 2007; Klein, Zatorre, Milner, and Zhao, 2001). For native speakers of Mandarin Chinese, tonal pattern is an integral part of each word they learn. They perceive and produce tones with reference to a system that defines not only what the tones *are* like but also what they are *not*, in relation to the other tones (Sun, 1997). However, for non-tonal speakers, such a

functional association between segmental structure and tonal contour is nonexistent (Sun, 1997; Tao, Beutner, and Bond, 2006; Tao and Guo, 2008; Wang *et al.*, 2003; Wang and Sereno, 1999). That is, L2 learners of Chinese with non-tonal L1 backgrounds such as English, Japanese, or Korean, do not have a phonological system that corresponds to the way native Chinese speakers perceive and produce the linguistically relevant information of lexical tones (Berkowitz and Strange, 2007; Hallé, Chang, and Best, 2004; Hao and de Jong, 2007).

Several cross-linguistic perception training studies have examined how speakers with tonal and non-tonal L1s differ in the learners' perceptual processing of tones (Tao and Guo, 2008; Wang *et al.*, 2003; Wang and Kuhl, 2003; Wang and Sereno, 1999). These studies show evidence supporting the hypothesis that tonally naïve adult learners could develop ad hoc phonetic categories for distinguishing two lexically contrastive tones although the studies also suggest that L2 learners were less categorical in their perceptual labeling compared to native speakers (Hao and de Jong, 2007; Wang *et al.*, 2003; Wang and Sereno, 1999). Additionally, many studies examined learners' tonal error patterns and attempted to identify the difficulty levels of acquiring different tones (Chen, 1997; Miracle, 1989; Shen, 1989; Winke, 2007). However, these studies yielded inconsistent findings as to which tone was the most difficult for L2 learners of Chinese to differentiate and to produce.

Despite the substantial amount of research on Chinese tones that has been undertaken, some fundamental issues need to be explicitly addressed. Most of the studies to date on Mandarin tones have focused on the effects of perception training with American learners of Chinese in laboratory settings (e.g. Tao and Guo, 2008; Wang *et al.*, 2001; 2003; Wang and Kuhl, 2003; Wang and Sereno, 1999). Although the effects of perception training on tone recognition have been well documented (Wang *et al.*, 2001; 2003; 2006), studies that examine the production

of tones have received relatively little attention (e.g. Chen, 1997; Shen, 1989; Winke, 2007). Additionally, very few studies have examined the contribution of tones to perceived degree of foreign accent, except Wayland (1997), who detailed the acoustic characteristics of tones in Thai and the contribution of Thai tone production to foreign accent. No known research thus far has examined the relationship between Mandarin tone production and native listeners' perceived degree of foreign accent and comprehensibility. To bridge the gap in the literature, this study aimed to investigate tone production and its association with perceived degree of foreign accent and comprehensibility among adult L2 learners of Chinese. Research questions guiding this study were:

1. What are the interrelationships among ratings of accentedness, comprehensibility, and tone production among L2 learners of Chinese?
2. To what extent do ratings of tone production contribute to degree of foreign accent?
3. Is there a difference among ratings of accentedness, comprehensibility, and tone production across L2 learners of Chinese with different L1 backgrounds?

Three hypotheses were formulated:

1. More heavily foreign-accented Chinese L2 speech is associated with a higher level of difficulty in comprehension.
2. Better production in Mandarin Chinese tones is associated with a lesser degree of foreign accent.
3. Better production in Mandarin Chinese tones is associated with a lower level of difficulty in comprehension.

Method

Participants

Participants were 30 adult L2 learners of Chinese who were enrolled in an intensive Chinese as a Second Language (CSL) program at a large university in China. They were 10 English, 12 Japanese, and eight Korean L1 speakers. The participants were at the intermediate-mid to advanced level of Chinese oral proficiency on the ACTFL scale assessed by an in-house Chinese proficiency test at the university where the study took place. They ranged in age from 20 to 46, with a mean of 29.4 years ($SD = 6.6$). The average age of the participants' AOL was 26.9 ($SD = 6.4$). The earliest AOL was 16. The participants' self-reported length of studying (LOS) Chinese was two and a half years ($SD = 0.9$). Details about the participants' personal backgrounds across the three language groups are provided in Table 1.

INSERT TABLE 1 APPROXIMATELY HERE

Raters

Two native speakers of Mandarin Chinese were recruited to rate the speech samples. They were postgraduate students who were enrolled in language-related programs at a large Midwestern university in the U.S. Both were teaching Chinese as a foreign language at the time of the study. The raters had an academic background in Chinese phonetics and phonology and had previous experience rating Chinese L2 speech.

Materials

Questionnaires A participant background questionnaire was administered. The questionnaire included items related to age, gender, learning experience, such as the total number of years spent studying Chinese, the amount of language use at home and at work, and types of previous Chinese instruction. A rater questionnaire was administered to obtain raters' background information and teaching experience.

Speech samples The 30 participants voluntarily took an in-house oral proficiency test administered by the Chinese language program where the study took place. The test is a semi-direct oral proficiency test, modeling the test format of the Chinese Simulated Oral Proficiency Interview (SOPI) (Stansfield, 1992; 1996). The test comprises 12 tasks and is 18 minutes in length.

For the purpose of the current study, a pilot study was first conducted to investigate the length of time native Chinese raters needed for making a consistent judgment on the participants' oral responses. It was found that raters were able to make a reasonably consistent judgment after they listened to different speech samples for a total length of approximately five minutes. In order to reduce the workload imposed on the raters, four tasks from each participant's test database were chosen for rating. The four tasks were approximately five minutes long in total. They were used for all the ratings of accentedness, comprehensibility, and tone production. The rating tasks differed in task type, difficulty, topic, and average length. The tasks were saved in an online rating system that was developed for the current project. The 30 participants were randomly presented with each task to avoid any ordering effect.

Rating scale There were three rating components: accentedness, comprehensibility, and tone production. Following Munro and Derwing (1995a), I used a 9-point scale for the ratings of accentedness (1= no accent, 9= extremely strong accent) and comprehensibility (1= very easy to understand, 9= extremely difficult or impossible to understand). I developed the rating scale for tone production (see Appendix), following the ACTFL guidelines for Chinese oral proficiency (ACTFL, 1987), and the rating scale development guidelines suggested by Luoma (2004). The scale descriptors were reviewed by two Chinese language teachers for clarity before they were pilot-tested. Results of the pilot study confirmed the use of the descriptors and the validity of the

5-point scale. Raters of this study also commented on the ease of using a 5-point scale to rate the participants' tone production during the training session.

It should be noted that the use of a rating scale to rate tone production assumes that tone production is a scalar construct (i.e. it exists on a continuum) rather than a binary (i.e. right or wrong) phenomenon. This concept of Chinese tones differs greatly from most previous perception training studies that adopted a binary definition of Chinese tones and employed different measures (e.g. discrimination or identification tasks) (e.g. Hallé *et al.*, 2004; Wang *et al.*, 2003; Wang and Sereno, 1999). The construct validity of the tone production rating scale will be discussed later in the paper.

Procedure

In order to reduce the variability of raters and to increase the reliability of the subjective ratings, a training session was conducted before the actual rating. Raters first filled out the background questionnaire at the beginning of the training session. The training was followed by the introduction of the purpose of the study and a detailed explanation of the rating scale. Then a set of responses was selected from the SOPI Rater Training Kit (Clark and Li, 1986) for the calibration exercise and was played to familiarize the raters with the rating scale. Raters were then taught how to interpret and use the rating scale and how to assign scores based on the performance of each task. The training session and calibration exercises lasted one hour. After the training, raters took a break and then rated the speech samples online individually, with additional breaks taken as needed. The rating took approximately three hours.

Results

Rater reliability was estimated using Cronbach's α . Inter-rater reliability was .81 for the ratings of accentedness, .82 for comprehensibility, and .88 for tone production. These estimates

were comparable to those reported in other related studies (e.g. Derwing, Thompson, and Munro, 2006) and were indicative of acceptable agreement between the two raters. Given the minor variability in rater agreement, the participants' final scores on the ratings that raters disagreed upon were derived from the average of the two raters' ratings across the four rating tasks. This was modeled after previous studies that determined final ratings using average scores across raters (Derwing and Munro, 1997; Munro and Derwing, 1998).

For each of the three measures (accentedness, comprehensibility, and tone production), I pooled the ratings on the four rating tasks over the two raters to compute a mean rating for each measure (see Table 2). The mean ratings of each measure were used in further inferential statistical analyses. The mean accentedness ratings for the 30 participants ranged from 1.8 to 7 with a group mean of 4.8 on the 9-point scale. The mean comprehensibility ratings ranged from 1.75 to 6, with a group mean of 3.4 on the 9-point scale. The mean tone production ratings ranged from 2.25 to 5, with a group mean of 3.2 on the 5-point scale. Although the Japanese participant's speech was perceived as the most heavily accented and least comprehensible, their tone production was rated the best ($M= 3.4$). On the contrary, the English participant's speech was rated the least foreign-accented and the easiest to understand, but their tone production was in the middle range with a mean of 3.2. The Korean speakers were perceived to have relatively heavy accent ($M= 5.0$) and received the lowest rating on tone production ($M= 2.8$).

INSERT TABLE 2 APPROXIMATELY HERE

To ensure that ratings of accentedness, comprehensibility, and tone production produced by the three participant groups were comparable, the participants' AOL and LOS were first compared using one-way ANOVA tests. No significant difference was found among the three

participant groups' AOL ($F(2, 27)= 1.041, p= .367$) and LOS ($F(2, 27)= .53, p=.949$). The statistics indicate that the three participant groups' backgrounds were relatively comparable.

Another set of one-way ANOVA tests was conducted to determine if ratings of accentedness, comprehensibility, and tone production differed across the three participant groups. Ratings of the three measures were entered as dependent variables and the participant's L1 was the independent variable. No significant difference was found on the ratings of accentedness ($F(2, 27)= 1.92, p= 0.16$), comprehensibility ($F(2, 27)= 2.02, p=.15$), and tone production ($F(2, 27)= 2.34, p =.11$) based on the L1 of the participants. Based on these results, the following analyses will focus on the participants as a whole.

In order to examine the interrelationships among accentedness, comprehensibility, and tone production, following Munro and Derwing (1995a), I plotted the distribution of the rating categories on all the rating tasks across the three measures. Figures 1 and 2 illustrate the percentage of the rating categories awarded by the raters for accentedness and comprehensibility, respectively. The accent ratings (Figure 1) are well distributed across categories 2 to 8. Twenty-seven percent of the ratings were given a rating of 7 or above, indicating a strong foreign accent. The comprehensibility ratings (Figure 2) show a different pattern. Nine percent of the rating tasks were awarded a rating of 1, indicating extremely easy to understand and 55 percent of the rating tasks received a rating of 3 or lower, suggesting that the speech samples were relatively easy to comprehend. In sharp contrast to accent ratings on the higher end, only 0.8 percent of the rating tasks were awarded a rating of 7 and no rating of 8 and 9 was observed on the comprehensibility ratings.

INSERT FIGURE 1 APPROXIMATELY HERE

INSERT FIGURE 2 APPROXIMATELY HERE

Figure 3 illustrates the distribution of tone production ratings. The tone production ratings are fairly normally distributed across categories 2 to 5. A very small percentage of rating tasks (1.7%) was rated 1, indicating unintelligible tone production. Almost half of the rating tasks (48.3%) received a rating of 3, indicating that these participants have not mastered the four tones; however, they could get the general meaning across. Additionally, 27.6 percent of the rating tasks received a rating of 4, and 7.3 percent received a rating of 5, which signifies a near native-like production in tones. Taken together, more than one third of the rating tasks were awarded a rating higher than the overall group meaning rating ($M= 3.2$).

INSERT FIGURE 3 APPROXIMATELY HERE

To further examine the interrelationships among accentedness, comprehensibility, and tone production, the mean ratings of these three sets of measures were subject to a Pearson product moment correlation analysis. Ratings of accentedness correlated highly with that of comprehensibility ($r= .913, p= .001$). However, accentedness ratings did not significantly correlate with ratings of tone production. On the other hand, comprehensibility ratings significantly correlated with ratings of tone production, although the strength of the correlation was moderate ($r= -.429, p= .018$). Finally, a simple linear regression analysis was conducted to examine the extent to which tone production contributed to perceived degree of foreign accent ($F(1, 28)=3.1, p= .088, R^2= .10$). The statistics indicated that tone production was not a significant predictor of the degree of foreign accent.

Discussion

In response to Research Question 1 that dealt with the interrelationships among ratings of foreign accent, comprehensibility, and tone production in the speech of L2 learners of Chinese, results of this study show that degree of foreign accent was highly correlated with perceived

comprehensibility, regardless of the participants' L1 backgrounds. However, accentedness ratings were not associated with ratings of tone production. The strong relationship between accentedness and comprehensibility supports the first hypothesis which claims that more heavily foreign-accented Chinese L2 speech is associated with a higher level of difficulty in comprehension. This result also corroborates findings of previous studies that indicate when L2 speech is perceived to be more heavily accented, it tends to be more difficult to understand (Burda and Hageman, 2005; Kennedy and Trofimovich, 2008; Munro and Derwing, 1995a; 1995b; 1998; 2001; Munro *et al.*, 2006). Furthermore, the analysis on the rating category distribution across rating tasks supports findings of Munro and Derwing (1995a), which suggest that raters tend to be harsher when they assign scores on the degree of foreign accent even though they consider the speech to be highly comprehensible. Additionally, although a strong correlation between accentedness and comprehensibility was observed, no significant correlation was found between accentedness and tone production. The result here does not support the second hypothesis which claimed that better tone production would be associated with a lesser degree of foreign accent.

Turning to the relationship between comprehensibility and tone production, the results reveal a different pattern. A significant correlation between ratings of comprehensibility and tone production was observed, suggesting that when a participant made more tonal errors, his/her speech was perceived to be more difficult to understand. This finding confirms the third hypothesis which claims that better tone production is associated with a lower level of difficulty in comprehension and implies that inaccurate production of tones presented a comprehension challenge for the raters. Nonetheless, this phenomenon is not surprising given that accurate production of tones is essential for conveying accurate information in Chinese speaker's speech.

The pedagogical implication from this finding is that, as Feng (2004) suggests, training in accurate production of tones is a priority in building a solid foundation in the speech of L2 learners of Chinese if their speech is to be precisely and accurately understood by listeners.

In response to Research Question 2 that asked the extent to which tone production contributes to degree of foreign accent, results of simple linear regression analysis indicated that the participants' tone production ratings were a poor reflection of perceived degree of foreign accent. However, this finding is counterintuitive. Because lexical tones have been widely shown to pose great acquisition difficulty for L2 learners of Chinese with non-tonal L1 backgrounds (Feng, 2004; Sun, 1997; Tao and Guo, 2008; Wang *et al.*, 2001; 2003), it was expected that tone production would correlate strongly with degree of foreign accent. The result here also differs from what Wayland (1997) found in her study that showed a strong relationship between accentedness and different contour tones produced by American L2 learners of Thai.

One possibility for the lack of significant predicting power of tone production for perceived degree of foreign accent is that the participants' segmental errors and other prosodic characteristics, such as intonation or stress pattern, along with fluency features, such as speech rate or pausing errors, may have outweighed the contribution of tone production on perceived degree of foreign accent. Many previous studies have documented the impact of different segmental and suprasegmental errors on perceived degree of foreign accent (Anderson-Hsieh *et al.*, 1992; Jilka, 2000; Munro, 1993; Munro and Derwing, 1998; 2001; Piske *et al.*, 2001; Trofimovich and Baker, 2006; van Els and de Bot, 1987). It is speculated that when the native listeners were listening to the speech samples, they may have first listened for the overall accent. When a strong accent was detected, their attention was directed to different properties of the speech that contributed to their perception of foreign accent (Wayland, 1997). These properties

may not have been the participants' production of tones per se. For the two raters in this current study, the participants' tone production may have played a less powerful role in causing their speech to sound deviant from native speaker norms.

Another possibility regarding the lack of predicting power of tone production for foreign accent pertains to the rating methods employed. This study utilized only one approach to assessing perceived degree of foreign accent and tone production, that is, the use of native speakers' subjective judgments. Raters' subjective judgments on degree of foreign accent could not reveal which specific acoustic characteristics of the participants' speech had caused them to perceive the speech as being foreign accented (Munro, 1993; Wayland, 1997). Thus, additional acoustic analysis that examines both segmental and suprasegmental parameters of the speech produced by L2 learners of Chinese and their contribution to degree of foreign accent is needed in future research. Such analysis will help determine which of these parameters influence perceived degree of foreign accent and supplement the subjective rating data (Munro, 1993; 1995; Munro and Derwing, 1998; Wayland, 1997; Wayland, Kaan, Bao, and Barkley, 2007).

Taken together, results of this study reveal that the difficulties raters had in understanding the participants' speech were reflected in their ratings of tone production and comprehensibility; that is, less accurate tone production was rated more difficult to understand. In addition, what was more difficult to understand was also rated as more accented as reflected in the strong relationship between ratings of accentedness and comprehensibility. However, better tone production was not associated with a lower degree of foreign accent and tone production did not contribute to perceived accentedness. Even though the participants' tone production was relatively good, which also indicates better intelligibility, raters still considered heavily accented speech more difficult to understand. The complex interrelationships among accentedness,

comprehensibility, and intelligibility support the results of several previous studies carried out by Derwing and Munro (Derwing and Munro, 1997; 2005; Derwing *et al.*, 1998; Munro and Derwing, 1995a; 1995b; 2001; Munro *et al.*, 2006) that provided evidence to show that while L2 speech may be intelligible, native listeners may still perceive foreign-accented L2 speech to be more difficult to understand.

Research Question 3 asked whether ratings of accentedness, comprehensibility, and tone production differed across the three participant groups. Although the mean ratings of the three measures varied across groups, the between-group mean rating differences were minor and not statistically significant. This result signifies that the participant's L1 background was not a robust factor that influenced the raters' judgments on foreign accent, comprehensibility, and tone production. The outcome suggests that properties of the participants' speech itself may have been a more influential determinant of the ratings compared to the potential effects of L1 backgrounds (Munro *et al.*, 2006). Nevertheless, due to the small sample size within each participant group, the results need to be interpreted with caution. The small sample size may have constrained the statistical power of the ANOVA tests and resulted in a non-significant effects of L1 backgrounds on the ratings of the three separate measures. Future research needs to include a larger sample size of participants with different L1s to test the findings derived here.

Although no significant difference on ratings of accentedness, comprehensibility, and tone production was found across the three participant groups, one noteworthy within-group observation deserves further discussion. The interrelationships among the three measures observed within the Japanese participant group show a baffling pattern. Despite the fact that the L1 Japanese speakers received the best ratings on tone production among the three groups, their speech was rated to be the most heavily accented and the most difficult to understand. The result

seems puzzling at least at first glance because, as discussed earlier, better production in tones was associated with better comprehension. It is not clear why the Japanese speakers' speech exhibited a counter relationship between comprehensibility and tone production. Apart from the possible segmental and prosodic errors the Japanese speakers may have made that led to the difficulty in comprehension, the problem may also arise due to negative L1 grammar interference and transfer (Gass and Selinker, 1994). The drastic differences in the syntactic structures of the Chinese and Japanese languages have been found to be a major source of variability in the interlanguage system of Japanese learners of Chinese (Takita, 2009; Yuan, 2007). Raters in this study may have noticed the non-native-like grammar in the Japanese participants' speech and perceived the speech to be more difficult to process although they may have considered that the speakers' tone production on the individual word or sentence level to be relatively accurate. While a thorough analysis of the differences in the phonetic systems and syntactic structures between the Chinese and Japanese languages is beyond the scope of this study, more fine-grained, cross-linguistic studies in this regard and a comprehensive investigation of the interlanguage systems of L2 learners of Chinese with different L1 backgrounds is highly desirable. Such investigations can provide more precise and in-depth understanding of the reason why the within and between group variability exists, as was found in this study.

The study presented here differs from many previous studies in several aspects. First of all, the study focuses on tone production of L2 learners of Chinese across three different L1s while most previous studies predominately examined American learners of Chinese (Feng, 2004; Hao and de Jong, 2007; Sun, 1997; Tao *et al.*, 2006; Tao and Guo, 2008; Wang *et al.*, 2003; Wang and Kuhl, 2003; Wang and Sereno, 1999; Winke, 2007). Secondly, the participants in this study were Chinese learners who studied in a CSL environment in China and were at

intermediate to advanced levels of study while the participants in most previous studies were beginning learners of Chinese studying in a foreign language environment such as the U.S. (e.g. Feng, 2004; Tao and Guo, 2008; Wang *et al.*, 2003; Wang and Sereno, 1999; Winke, 2007). Thirdly, this study used longer, spontaneously produced speech responses to examine the participants' degree of foreign accent, perceived comprehensibility, and quality of tone production, as opposed to word- or sentence-level oral stimuli evaluated in many previous studies (Wang *et al.*, 2001; 2003; Wang and Kuhl, 2003; Wang and Sereno, 1999). Thus, findings of this study are more generalizable to naturally produced L2 speech in Chinese. Finally, this study employed a tone production rating scale empirically developed by the researcher and thus provided a new measure of tone production for future validation study in the assessment of Chinese oral proficiency.

Several limitations need to be addressed. First of all, no native speakers' speech samples were included in this study and thus left unanswered the extent to which the participants' speech deviated from the native speaker norms. Secondly, only 30 participants' oral responses were used for evaluation. The small sample size may have constrained the statistical power of the inferential statistical analysis. In addition, only two native raters were recruited to rate the speech samples. The interpretation of the results was solely based on the two raters' subjective judgments. In order to rectify these methodological issues, the next step of this study will include gathering more speech samples from L2 learners of Chinese as well as native speakers. Furthermore, more native raters will be recruited to evaluate the speech samples. The third limitation pertains to the reliability regarding the use of the tone production rating scale. The reliability of the tone production ratings hinges, to some extent, on the extent to which raters could rate tone production consistently and accurately, regardless of segmental interference. It

can be argued that the participants' tone production ratings might not have been rated as highly if an objective acoustic analysis had been employed. This point brings into question the conceptualization of tone production as a continuous construct. It may be the case that tone production is a binary phenomenon as assumed in most previous perception training studies (Wang *et al.*, 2003; Wang and Sereno, 1999). This issue points to the construct definition of Chinese tones and needs to be further explored in future research.

Conclusion

This study investigated the interrelationships among accentedness, comprehensibility, and Mandarin tone production. Results of this study provide evidence supporting the argument that more heavily accented speech is more difficult for native listeners to understand (Derwing and Munro, 1997; 2005; Derwing *et al.*, 1998; Munro and Derwing, 1995a; 1995b; 2001; Munro *et al.*, 2006; Trofimovich and Baker, 2006) and better tone production is easier to comprehend. However, the L2 learner's L1 background was not found to be an influential factor in the raters' judgments of accentedness, comprehensibility, and tone production.

Since tone production did not accurately reflect the extent to which the L2 speech in Mandarin Chinese was perceived accented, this finding raises the question as to what factors contribute to the perception of accentedness among native Chinese listeners. To explore the exact nature of perceived foreign accent in the speech of L2 learners of Chinese, future research needs to investigate segmental and other suprasegmental aspects of L2 speech such as stress pattern, or intonation, and how they impact perceived degree of foreign accent (Anderson-Hsieh *et al.*, 1992; Derwing *et al.*, 1998; Munro, 1993; Trofimovich and Baker, 2006). Evidence along these lines will further advance our understanding of the interplay between tone production and perceived foreign accent. Additionally, more research is called for to investigate the construct of tone

production as a continuous as opposed to a dichotomous entity of Chinese speech. Findings of such investigations can better inform the assessment of Chinese oral proficiency in the future.

References

- ACTFL (1987). ACTFL Chinese proficiency guidelines. *Foreign Language Annals*, 20, 471-487.
- Anderson-Hsieh, J., Johnson, R., and Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42, 529-555.
- Bent, T., and Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, 114, 1600-1610.
- Berkowitz, S., and Strange, W. (2007). Discrimination of lexical tone contrasts in disyllabic nonsense pairs by adult speakers of Mandarin and of English. *Journal of the Acoustical Society of America*, 3021.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133-159). Mahwah, NJ: Erlbaum.
- Brown-Schmidt, S., and Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistics Research*, 33, 103-135.
- Burda, A. N., and Hageman, C. F. (2005). Perception of accented speech by residents in assisted-living facilities. *Journal of Medical Speech-Language Pathology*, 13, 7-14.
- Chao, Y. R. (1948). *Mandarin primer: an intensive course in spoken Chinese*. Cambridge, MA: Harvard University Press.
- Chen, Q. (1997). Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association*, 32, 21-39.

- Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics*, 20, 289-301.
- Clark, J.L.D., and Li, Y.C. (1986). *Development, validation, and dissemination of a proficiency-based test of speaking ability in Chinese and an associated assessment model for other less commonly taught languages*. Washington, DC: Center for Applied Linguistics.
- Derwing, T. M., and Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 20, 1-16.
- Derwing, T. M., and Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39, 379-397.
- Derwing, T. M., Munro, M. J., and Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48, 383-410.
- Derwing, T. M., Thompson, R. I., and Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, 34, 183-193.
- Feng, S. (2004). Prosodic structure and its implications for teaching Chinese as a second language. *Journal of the Chinese Language Teachers Association*, 39, 1-24.
- Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 84, 70-79.
- Flege, J. E., and Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, 91, 370-389.
- Flege, J. E., Munro, M. J., and MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 106, 2973-2987.

- Flege, J. E., Yeni-Komshian, G. H., and Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78-104.
- Fu, Q.-J., and Zeng, F.-G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing*, 5, 45-57.
- Gass, S., and Selinker, L. (1994). *Language transfer in language learning*. Philadelphia: Benjamins Publishing Co.
- Gass, S., and Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, 34, 65-89.
- Hallé, P. A., Chang, Y.-C., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32, 395-421.
- Hao, Y.-C., and de Jong, K. (2007). The categorical nature of tones and consonants: evidence from second language perception and production. *Journal of the Acoustical Society of America*, 122, 3018.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent: identifying intonational deviations by means of F0 generation and resynthesis*. Unpublished doctoral dissertation, University of Stuttgart, Stuttgart.
- Kennedy, S., and Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: the role of listener experience and semantic context. *The Canadian Modern Language Review*, 63, 459-489.

- Klein, D., Zatorre, R. J., Milner, B., and Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*, 13, 646-653.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Miracle, C. (1989). Tone production of American students of Chinese: a preliminary acoustic study. *Journal of the Chinese Language Teachers Association*, 24, 49-65.
- Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36, 39-66.
- Munro, M. J. (1995). Nonsegmental factors in foreign accent: ratings of filtered speech. *Studies in Second Language Acquisition*, 17, 17-34.
- Munro, M. J., and Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 41, 73-97.
- Munro, M. J., and Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38, 289-306.
- Munro, M. J., and Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, 48, 159-182.
- Munro, M. J., and Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23, 451-468.
- Munro, M. J., Derwing, T. M., and Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111-131.
- Piske, T., MacKay, I. R. A., and Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.

- Rasier, L., and Hiligsmann, P. (2007). Prosodic transfer from L1 to L2: Theoretical and methodological issues. *Nouveaux cahiers de linguistique française*, 28, 41-66.
- Riney, T. J., and Flege, J. E. (1998). Changes over time in global foreign accent and liquid indetifiability and accuracy. *Studies in Second Language Acquisition*, 20, 213-244.
- Shen, X. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, 24, 27-47.
- Shen, X. (1990). *The prosody of Mandarin Chinese*. Berkeley, CA: University of California Press.
- Stansfield, C.W. (1992). *The Preliminary Chinese Proficiency Test (Pre-CPT): Development, scaling and equating to the Chinese Proficiency Test (CPT)*. Washington, DC: Center for Applied Linguistics.
- Stansfield, C.W. (1996). *Test development handbook: Simulated oral proficiency interview*. Washington DC: Center for Applied Linguistics.
- Sun, S. H. (1997). *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese*. Honolulu, HI: University of Hawaii Press.
- Takita, K. (2009). If Chinese is head-initial, Japanese cannot be. *Journal of East Asian Linguistics*, 18, 41-61.
- Tao, L., Beutner, M., and Bond, Z. (2006). Speech recognition technology in the instruction of Mandarin Chinese. *Journal of the Chinese Language Teachers Association*, 41, 57-88.
- Tao, L., and Guo, L. (2008). Learning Chinese tones: A development account. *Journal of the Chinese Language Teachers Association*, 43, 17-46.
- Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, 41, 177-204.

- Trofimovich, P., and Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28, 1-30.
- van Els, T., and de Bot, K. (1987). The role of intonation in foreign accent. *Modern Language Journal*, 71, 147-155.
- Wang, Y., Jongman, A., and Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American learners. *Brain and Language*, 78, 332-348.
- Wang, Y., Jongman, A., and Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, 113, 1033-1043.
- Wang, Y., Jongman, A., and Sereno, J. A. (2006). L2 acquisition and processing of Mandarin tone. In P. Li, L. Tan, E. Bates, and O. J. L. Tzeng (Eds.), *Handbook of east Asian psycholinguistics* (pp.250-256). Cambridge: Cambridge University Press.
- Wang, Y., and Kuhl, P. K. (2003). Evaluating the "critical period" hypothesis: perceptual learning of Mandarin tones in American adults and American children at 6, 10 and 14 years of age [online]. Available:
http://ilabs.washington.edu/kuhl/pdf/wang_kuhl_2003.pdf
- Wang, Y., and Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106, 3649-3658.
- Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics*, 18, 345-373.

- Wayland, R., Kaan, E., Bao, M., and Barkley, C. (2007). Effects of native language and training on speaker normalization in lexical tone perception. *Journal of the Acoustical Society of America*, 122, 2997.
- White, C. M. (1981). Tonal pronunciation errors and interference from English intonation. *Journal of the Chinese Language Teachers Association*, 16, 27-56.
- Winke, P. (2007). Turning into tones: The effect of L1 background on L2 Chinese learners' tonal production. *Journal of the Chinese Language Teachers Association*, 42, 21-55.
- Ye, Y., and Connie, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14, 609-630.
- Yuan, B. (2007). Japanese speakers' second language Chinese wh-questions: A lexical morphological feature deficit account. *Second Language Research*, 23, 329-357.

Appendix

Tone production rating scale

TONE PRODUCTION	
5	Rarely mispronounces any tones in words and sentences; able to speak with correct tone Sandi spontaneously.
4	Tone production is clear, occasionally mispronounces tones; has mastered all tones. Inaccurate tone production does not interfere with meaning.
3	Tone production is not always correct, but the meaning can generally be understood. Often mispronounces unfamiliar words; has not mastered all tones.
2	Frequently mispronounces tones. Inaccurate tone production impedes meaning. Difficult to understand even with concentrated listening.
1	Hasn't mastered accurate tone production. Mispronounces most tones in word and sentence levels. Tone production is not intelligible.

Table 1

Participants background information

	All	English	Japanese	Korean
Gender	11 male 19 female	3 male 7 female	5 male 7 female	3 male 5 female
AOL	$M= 26.9$ ($SD=6.4$)	$M= 29.3$ ($SD= 7.5$)	$M= 25.4$ ($SD= 6.4$)	$M= 26.2$ ($SD= 4.9$)
LOS	$M= 2.5\text{yrs}$ ($SD=0.9$)	$M= 2.5\text{ yrs}$ ($SD = 1.2$)	$M= 2.5\text{ yrs}$ ($SD= 0.8$)	$M= 2.4\text{ yrs}$ ($SD= 0.7$)

Table 2

Mean ratings of accentedness, comprehensibility, and tone production

	All (<i>N</i> = 30)	English (<i>N</i> = 10)	Japanese (<i>N</i> = 12)	Korean (<i>N</i> = 8)
Accentedness	4.8 (<i>SD</i> =1.1)	4.3 (<i>SD</i> =1.4)	5.2 (<i>SD</i> =.9)	5.0 (<i>SD</i> = .7)
Comprehensibility	3.4 (<i>SD</i> = 1.1)	2.9 (<i>SD</i> =1.2)	3.8 (<i>SD</i> =1.1)	3.5 (<i>SD</i> =.9)
Tone production	3.2 (<i>SD</i> = .6)	3.2 (<i>SD</i> =.7)	3.4 (<i>SD</i> =.6)	2.8 (<i>SD</i> =.3)

Figure 1

Distribution of accentedness ratings (1= no accent, 9=extremely strong accent)

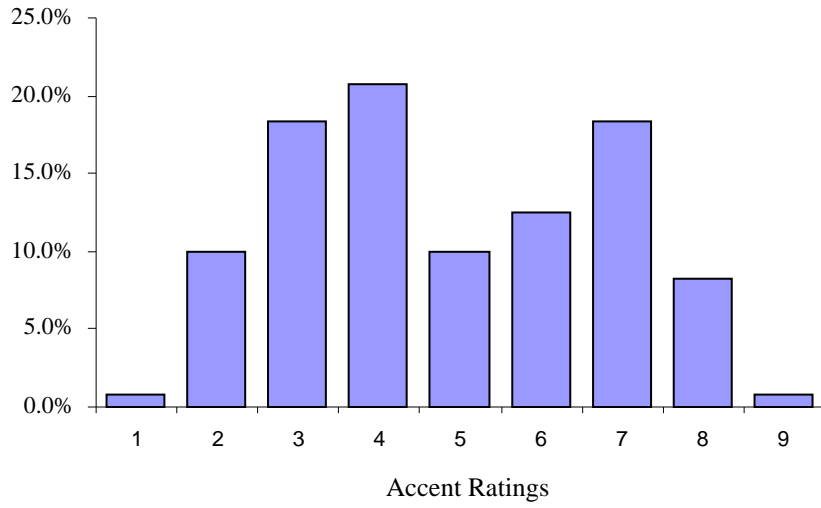


Figure 2

Distribution of comprehensibility ratings (1= extremely easy to understand; 9= impossible to understand)

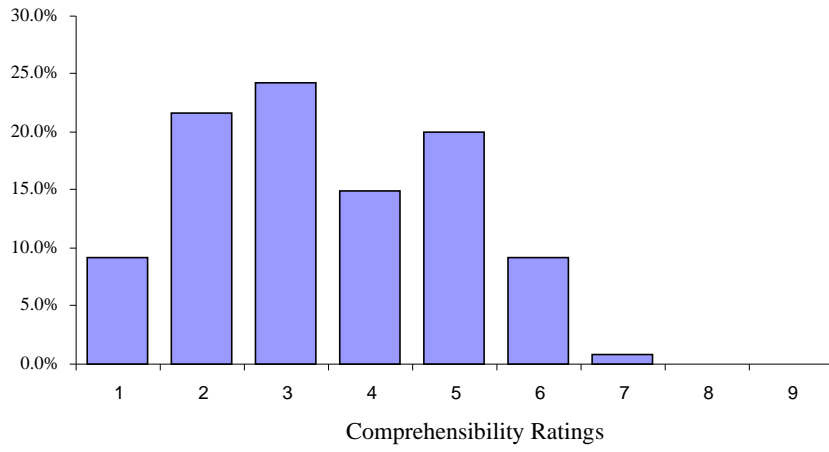
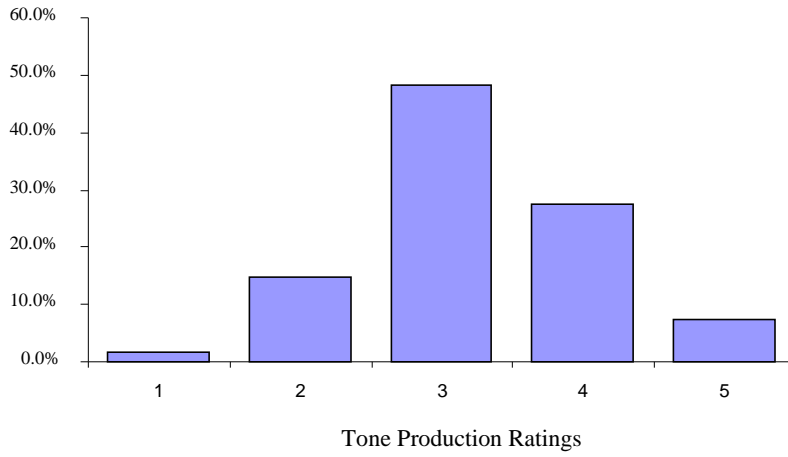


Figure 3

Distribution of tone production ratings (1= tone production unintelligible; 5=near native-like tone production)



Interview with Dr. Robert DeKeyser

Interviewed by Jimin Kahng
PhD Student, Second Language Studies Program
Michigan State University
kahngji@msu.edu

I would like to thank Robert DeKeyser, Professor of Second Language Acquisition, University of Maryland, for agreeing to be interviewed for these working papers. Dr. DeKeyser's research interests include SLA, interlanguage variability, individual differences in SLA, aptitude-treatment interaction, cognitive psychology of language acquisition, and age effects in SLA. He came to Michigan State University to give a plenary talk at the Second Language Research Forum 2009. The interview was conducted by Jimin Kahng on November 1, 2009. For more information about Dr. DeKeyser, please visit his faculty page: <http://www.languages.umd.edu/SLAA/rdk.html>.

Jimin: How did you come to be interested in second language acquisition?

Dr. DeKeyser: Well, living in Belgium, I guess, it's easy to do that because we had French as a second language in grade school, and at least in our high school track we all had six languages; two classical, Latin and Ancient Greek, two languages spoken in the country, French and Dutch, and English and a bit of German. So if you're learning languages all the time, you're interested in the process, of course. In a country where so many languages are taught, you see a lot of people studying language and literature at the university and a lot of language teachers and that was my plan, too. Then I decided I wanted to learn to do research alongside teaching and that's how this all started. (laugh)

Jimin: Moving on to some of your research, let's talk about implicit and explicit learning. Could you explain the relationship between the two types of learning and the role of practice?

Dr. DeKeyser: There are slightly different definitions depending on who you read, but I think my definition would be something like this: implicit learning is learning without awareness of what you're learning and explicit learning is learning with awareness of what you are learning.

If you believe in an important role for implicit learning, then practice basically means communication and from lots of communication, proficiency should result. If you think explicit learning is very important, then practice means any way of using the initial explicit knowledge you learned, using that to generate more proceduralized, automatized, and maybe eventually implicit knowledge. Whether knowledge eventually becomes implicit or not, I think few people really care about that from a practical point of view. The important thing is that you can automatize. Of course, the classical problem is that at the beginning you learn the rules and you may know them perfectly and be able to use them on a fill-in-the-blanks test, but using them for speaking is different, so the role of practice is gradually transforming that knowledge, using one kind of

knowledge to create another kind, which is probably the better metaphor, so that eventually you can speak fluently on the basis of what originally you were taught in the form of explicit rules.

Jimin: There are researchers who argue that there's no interface between implicit and explicit knowledge. I was reading a part of this book (showing the book *A Neurolinguistic Theory of Bilingualism* by Paradis) and...

Dr. DeKeyser: Oh, Michel Paradis, yes. He has a newer book by the way, by the same publisher, which is called *Declarative and Procedural Determinants of Second Languages*—I certainly read the chapter on age and agree completely with what he said about age. I haven't read the whole book yet but I do know that for a long time his point of view has been that you cannot turn one kind of knowledge into another. I agree with this point of view; one type of knowledge is CAUSAL in the development of the other.

I remember very well a couple of years ago, there was a symposium at Georgetown University, which is almost our neighbor, and Paradis and Ullman very strongly argued that you cannot turn explicit knowledge into implicit or, to use Ullman's term, "declarative" into procedural because anatomically and physiologically there is no neurological path for this process. But my point is that the role is causal and indirect in the sense that you can engage in certain tasks because of your explicit knowledge and then engaging in these tasks gradually builds up implicit knowledge. So it's not like this all happens inside the brain going from one place to another. No. It's really the tasks that explicit knowledge allows you to engage in, which then allows you to learn implicit knowledge from.

This is very ironic because, I'm sure this is something you won't know, nobody

knows this, if there is one person that I should credit for getting me interested in psycholinguistics, then it is Michel Paradis. Thirty years ago, when I was an undergraduate in Belgium, Michel Paradis on a rainy December afternoon gave a talk to the undergraduates about psycholinguistics and neurolinguistics. I also met with him briefly during or after graduate school in his office in Montreal, and of course, he's done a lot of very interesting research. So it's sort of ironic that after, in a way, being inspired by him to get into this profession, I end up disagreeing on a fairly important point. However, our disagreements are not that complete because, as I said, I really think it's due to misleading phraseology that people seem to disagree so much. "Turning into" is not the right word. It's "causal effect of one leading to the other". There are people that would deny that causal effect anyway. But I would put it as follows "what do we have most evidence for?" People learning knowledge explicitly and then using that to develop fluency, or people learning implicitly from scratch? What do we have most evidence for? Well, I would say, for the former case, we have evidence from millions of people; for the other case, we have evidence from zero people. (laugh)

Jimin: Could you briefly introduce your current projects?

Dr. DeKeyser: I have several things that I need to write up but what I'm most interested in at this point is designing a new study on age effects because it's a very controversial topic, very poorly understood and I would like to do something rather radically new methodologically speaking, not only on a larger scale because all the studies suffer from very small sample sizes but also with better selected people who can be argued to have had very good input,

people who have had minimal contact with their native language. Of course the combination of a large sample with people who have hardly any contact with their native language which is very hard find because if you're in a place with a lot of nonnative speakers that have the same native language, then they speak each other's language. And if you find a place where there is somebody who's isolated, then that's a good example of a good context for acquisition but then we have only one person and, of course, you need a lot more. So I'm trying to think of ways of finding a compromise between these two problems.

Jimin: How can Second Language Studies serve the community at large? What do you think is the role of SLS?

Dr. DeKeyser: We are more applied than most areas of linguistics. A lot of what we do can and should have practical implications for how people teach, for how people put curricula together and even for broader questions like at what age should you start teaching or how should you adapt teaching to age and so on.

The role here for research actually specifically could be geared a little bit more towards the underserved populations right now because we tend to always think of people taking foreign languages in high school and college, or people who are in an English language institute as graduate students but there are many immigrants in a lot of countries. That population tends to be underserved because they are not required to take language classes and, there is often not much funding for language classes. We don't exactly know how to deal with these very diverse populations and so on. So I think an important role for research is to look at populations with little education and who, of course, need English just like anybody else and to see how teaching

methods can be adopted, because even though I'm a great believer in explicit learning and explicit knowledge, I realize, of course, that you cannot teach lots of grammar rules to people with very limited time and very limited educational background, and that's sometimes you'll need a shortcut in the sense of teaching more lexical knowledge, more collocations, more pragmatics, and so on. You cannot do everything in a short period of time. Sometimes you have to do something radically different. There's a difference between ideal and what is the best you can do in many circumstances.

Jimin: What do you think are challenges the field of SLS is facing?

Dr. DeKeyser: I don't know where to start on this point because there are so many. One challenge is the misinterpretation of our research: for instance, even though myself, I strongly believe in age effects, I do NOT think that they imply you have to start teaching children at an early age. I'm very worried that in a number of countries in Europe and Asia, people are going to start teaching earlier and earlier and will then be very disappointed that they don't get better results and blame it on us, the researchers who told them to start early. Well, no, no, no. I'm not telling them to start early, I'm saying there are strong age effects with immigrants but what you find in the classroom is different. You cannot transfer those conclusions from immigrants to the classroom. So that's one problem, overgeneralization of our own research.

The second problem in the language teaching field is the fact that so often budgets are cut starting from language teaching in particular. Language teaching is expensive the way it is now. You can give an introductory physics lecture to 200, 300 people, it doesn't matter but, we all know

you cannot teach a language to a group of 200, 300 people. You have to have small sections, really 20 is already big and that makes instructions more expensive. In these days of budget cuts many universities are putting pressure on language departments to make their teaching almost entirely computerized and to save money that way. First of all, I'm not so convinced that they'll all save a lot of money because developing curricula and buying computers are not exactly free. On top of that, if the goal is that people develop communicative skills in a second language, how far can we go with all this computer business? If there is one thing computers are still poor at, it is teaching a language. Clearly you can replace the teacher or the textbook with a computer for the most extreme forms of declarative knowledge that I was talking about. You can provide grammar rules, simple exercises, word lists and all of that on the computer, but when it comes to understanding what a student is trying to say in a second language and giving feedback on that, there is no computer program on earth that can do that. So that is the limitation that people outside of the profession don't often realize. And it's enormously ironic from my point of view that now we have finally all learned, when I say "all", I mean not just happy few in academia in a couple of countries but worldwide a lot of language teachers know teaching should be communicative, it is very ironic that at this point when we realize how important communicative language teaching is we're switching to computers, which are so bad at this communicative aspect, only good at presenting the structural aspect. So that's another challenge.

Jimin: What kind of advice do you have for future second language researchers and teachers?

Dr. DeKeyser: The first thing any aspiring academics should know is that you'll always have to work hard. It's not going to be a forty-hour-per-week job or a 9 to 5 job. Most academics I know work just about every evening and just about every weekend, so you have to work hard and you'll never get rich. But the interesting part of the job is, of course, that it's hardly ever routine. Research, of course, by definition is not a routine, otherwise it wouldn't be research. The teaching is not very routine either because you always have new students, somewhat new subject matters, so it is from that point of view, a challenging but very interesting profession. And you have to have a level of creativity and also the willingness to go through sometimes tedious boring detail to make things work as a researcher.

Dr. DeKeyser participated in this interview with enthusiasm and sincerity by sharing his ideas on various issues in SLA and by providing thoughtful and practical advice for future second language researchers and teachers. I believe his interview offers an opportunity for us to see a bigger picture and think about the role of researchers and teachers in SLA.

(Submission received 28 May 2010)

(Revision received 28 May 2010)

(Revision accepted 15 July 2010)

Interview with Dr. Roy Lyster

Interviewed by Seongmee Ahn
PhD Student, Second Language Studies Program
Michigan State University
ahnseon2@msu.edu

Thank you to Dr. Roy Lyster for accepting our invitation for SLS Working Papers. Dr. Lyster is Professor of Second Language Education, Department of Integrated Studies in Education, at McGill University. His research interests include second language acquisition in classroom settings in general, with a specific focus on immersion and content-based classrooms, more specifically, teacher-student interaction, form-focused instruction, and corrective feedback. He came to Michigan State University to give a talk at Second Language Research Forum. The interview was conducted by Seongmee Ahn on November 1, 2009. For more information about Dr. Lyster, please visit his faculty page: <http://people.mcgill.ca/roy.lyster/>.

Seongmee: How did you become interested in second language research?

Dr. Lyster: I've always been interested in languages and am a second language learner myself. I grew up in Saskatchewan, Canada. It's monolingual English, but I became interested in French. I studied with the audio-lingual method in elementary school and then, in high school, I became really good at it. I went on a student exchange in Quebec City where I was fascinated to learn about a vibrant culture other than Anglo-American culture. I became committed to becoming fluent in French. My bachelor's degree is in French language and literature. I got a scholarship from the French government to do an MA in French literature, and I lived in Paris for three years. When I came back to Canada in the early eighties, I settled in Toronto where immersion was a really big thing. Everybody said, "If you could get your teaching degree, you could start teaching in immersion." So, I did that and got a job in

French immersion, which I did for almost ten years. During that time, I started my master's degree at OISE. I worked with people, like Birgit Harley, Merrill Swain, Sharon Lapkin, and Jim Cummins. I loved doing my master's, because I found it very interesting to be a teacher in immersion. I did night courses for my master's, which helped me reflect on my teaching and understand theoretically better why my students spoke the way they did. Through my PhD dissertation, I became interested in research. All sorts of questions started to arise concerning immersion like, "How do you teach a language at the same time as teaching content?" Those questions helped me stumble across corrective feedback because it seems to be quite central. We used to think that feedback was just kind of the tip of the iceberg or icing on the cake, but I actually think it's fundamental.

Seongmee: Why are you interested in classroom-based research?

Dr. Lyster: I was a teacher for ten years, but I had many questions about my teaching effectiveness. There wasn't much guidance about how to teach language in an immersion context so I became interested in different models of instructional intervention. Now I work at a faculty of education, doing teacher training, so I am still interested in pedagogy and working with teachers in classrooms. Lab studies have really fallen short of giving sufficient information about language learning in classroom settings.

Seongmee: What advice would you give to language teachers about how best to provide corrective feedback in their language classrooms?

Dr. Lyster: Feel confident when giving corrective feedback. Don't worry about its effectiveness as the evidence is clear that providing feedback in classrooms is more effective than not. Many teachers feel that students don't want feedback because it hurts students' feelings, but a lot of research shows that students want to receive feedback. Create contexts that are meaningful enough for the feedback to be useful and effective. In terms of feedback types, with younger learners, don't use much explicit, metalinguistic explanation, although older learners might like that. Research shows that prompting is probably more effective, but you can only prompt learners in what they've already begun to acquire. So, you have to make online decisions about the level that your students are at. If you know that your questions are beyond your students' interlanguage development, then it's a perfect time to use recasts, but don't overuse recasts. It takes less time to prompt than people suggest. It can be very expeditious to give short prompts, trying to push learners to self-repair.

Seongmee: Would you have any advice for graduate students to become a competent SLS researcher?

Dr. Lyster: Good students have to be well-read and delve into all sorts of theories and theoretical perspectives. You need to do empirical research and draw on a range of theoretical orientations to help you interpret and understand your results. Too many researchers are driven by one theoretical orientation and design their study to justify it. You need to set up a study that allows you to keep the possibilities open and draw on theory afterwards to interpret data.

Seongmee: Could you please tell us about your current projects or initiatives?

Dr. Lyster: I'm in-between projects. I've completed a bilingual read-aloud research project with Laura Collins in immersion classrooms in Montreal. It's being published in *Language Awareness*. I'm trying to continue in that classroom-based vein. I'm working on a pedagogical project with teachers in a French immersion school. I'm trying to have the English and French teachers of the same students collaborate, which they normally don't do. I'm convinced that, for content-based language instruction to be effective, the teachers of both languages need to work together. In feedback, I work more and more with my students. I'm hoping to work more with Kazuya Saito who's interested in feedback and its role in pronunciation teaching. I finished an interesting study with Yingli Yang. It will be coming out in *Studies in Second Language Acquisition* where we did a feedback study, a classroom intervention study, basically comparing no feedback, prompts, and recasts, but in a Chinese EFL context. I'm interested in doing more research in a foreign language context, not just immersion settings in Canada. I'm

counting on my international students to take me in new directions.

Seongmee: Would you tell us about some research gaps or future directions, specifically in corrective feedback, or generally in second language research?

Dr. Lyster: There's still more research to be done on corrective feedback. But, in SLS research, I'm more interested in going beyond the input-output interaction model. I'm interested in feedback, practice, and the role of output, which is insufficiently explored. When Swain proposed the output hypothesis, research moved more towards meta-talk and the use of language to talk about a language. It's interesting to come back to the role of output in developing fluency through practice. So, I'm interested in how feedback and practice can work in tandem effectively in classroom settings and trying to move away from the idea that has dominated SLS research that learning should be implicit. I don't think there is sufficient evidence that we learn languages best implicitly, certainly not in classroom situations. I don't know that explicit instruction is the best, but I would like us to move away from the idea of demonstrating through research that focus-on-form has to be implicit. I hope we can look at the role of practice and output in the context of communicative and content-based instruction. Content-based instruction is the future of language teaching, which is what I'd like to see for classroom SLS research.

Seongmee: Do you have additional comments you would like to make before we end our interview?

Dr. Lyster: I haven't been in SLRF for quite a while. I was here for SLRF in 1997 and then at Columbia University in 2005. Anyway, I really enjoyed the conference,

and I think that it seems like an interesting student body. I loved the venue. This is a great place for a conference. I think that the students did a really good job. So, I would just like to say, "Congratulations!"

Seongmee: Thank you for your kind words and agreeing to interview for *SLS Working Papers*. It was a great pleasure to talk with you.

Dr. Lyster: Thank you. It was fun for me.

(Submission received 18 May 2010)

(Revision received 18 May 2010)

(Revision accepted 29 May 2010)

Interview with Dr. Jenefer Philp

Interviewed by *Sheila Wheeler*
Master Student, TESOL Program
Michigan State University
wheel173@msu.edu

Thank you to Dr. Jenefer Philp, Senior Lecturer of Applied Language Studies and Linguistics, The University of Auckland, for granting this interview. Dr. Philp's research focuses on second language acquisition and individual differences, interaction, and oral corrective feedback, as well as on child second language acquisition. She came to Michigan State University to present at the Second Language Research Forum (SLRF) 2009, which was held October 29th through November 1st. For more information about Dr. Philp, please see her faculty page: <http://artsfaculty.auckland.ac.nz/staff/?UPI=jphi048>. This interview was arranged by Sheila Wheeler and conducted on October 31st.

Sheila: Can you tell me about how you first got involved in Second Language Studies?

Dr. Philp: Well I used to teach French in high school, retrained for ESL and then worked in Indonesia for three years as an ESL teacher. I became involved in a teacher training project, a World Bank project there. And while I was teaching there I really felt that I couldn't teach well if I didn't know how people learned languages, and so after three years in Indonesia I went back to do my master's to find out how people learn languages. So I did my M.Phil in applied linguistics at the University of Sydney - I was really lucky that Cathy Doughty was there, and I was able to do my dissertation with Cathy on peer interaction in the foreign language classroom. The University of Sydney hosted the first Pac-SLRF (Pacific SLRF). It was great to hear the plenary speakers talk about interaction and SLA. I can really remember being there and people talking and they said "What do you think Mike?" and "What do you think Teri?" - all these people whose articles I'd read were in the room, and Sue Gass was there. You know it was so exciting to actually meet

them and talk with them. I got really interested in it. Although it was always my plan go back to Indonesia or somewhere else and continue teaching, I ended up working in a center for research on Second Language Acquisition (LARC), under Manfred Pienemann. Alison Mackey was working there too. So, you know, I was kind of involved in this world of SLA research and I found it really interesting. But originally it came from a desire to be able to teach languages better.

Sheila: Have you done any teaching since you went to school to get your master's?

Dr. Philp: Oh yeah, because when I was doing my master's I worked part time teaching English and also with the PhD I did ESL teaching at the same time. I'm now at the University of Auckland in New Zealand, as well as lecturing in SLA and other areas I've taught English for Academic Purposes as part of the program offered by our department.

Sheila: What kind of advice do you have for current students, either MA TESOL or SLS

PhD students, but specifically students who want to go on to teach?

Dr. Philp: I think being at university and doing a master's or PhD is really an opportunity to think about teaching, and think about what's behind how people learn languages ... to think about how does that relate to teaching, and ways of teaching. I teach a course on individual differences and a lot of the people that I teach are teachers, or people wanting to be teachers, they often remark it's really good to think about the learners and where they're coming from and what's helpful for them, and how they're all different. You know I think it's a time that you can actually reflect on the whole learning process, because when you're teaching you're so busy teaching you don't really have time to read up and to think about it. So make the most of the time that you've got to do that and once you begin teaching, think about ways to keep those links and to keep reading, coming to conferences, talking to people, or have research done in your classroom, or do action research yourself, publish in journals – these are all good ways of linking research and teaching. I think if you're a reflective teacher that's very helpful. You're thinking about your teaching all the time and how you can improve, and keeping up on new developments. If you haven't taught before I think it's really good if you can get opportunities for voluntary or paid work teaching and observing classrooms, because that helps you understand better the whole process. I think if you just go straight through your degrees without teaching, you know, you can read it all but it doesn't mean as much until you relate it to actual teaching. It's good to complement studying with observing or teaching, just like when you're teaching it's good to keep reading.

Sheila: What about the PhD students?

Dr. Philp: It depends on what area you want to work in. Some people go into research or go into areas like the government departments, for example government departments in education would maybe employ people to advise teachers or to be involved in testing, reporting, working out policies or materials development. So I think if you're interested in that area, it's good to have teaching experience. Sometimes even going away for a year, going abroad for a year, you know volunteering can open possibilities. I'm not really familiar with the American situation, in New Zealand where people with higher degrees find work is in schools, in government or private enterprises that adjunct to educational institutions, - in materials development, teacher training, teacher development, language testing - it just depends what your expertise is, your research and experience, as to where you want to go with it.

Sheila: Can you tell me a little bit about what you're working on currently?

Dr. Philp: I have two areas of interest at the moment, and one has to do with child language acquisition. My broad area's on interaction in SLA. So one area is on child second language acquisition, and differences between age groups. In some of my research I investigate how 5 to 7 year olds are different from 11 to 12 year olds are different from adolescents: how interaction is different for those different ages; how learning processes might be different for those different ages; how that could affect the way we teach. So part of the research is related to understanding SLA and part of the research is related to teaching/instructed SLA. So that's one area and another area that I'm interested in at the moment is peer interaction in second language classrooms, because I think in classrooms learners spend a lot of time talking to other learners, rather

than with the teacher. That's an area that's very messy to research, that I'm just starting to learn more about.

(Submission received 23 May 2010)

(Revision received 23 May 2010)

(Revision accepted 29 May 2010)

Sheila: Do you sense that there may be a risk with two non-native speakers, that they may incorporate each other's non-nativelike utterances?

Dr. Philp: Yeah, sure. I think it depends a lot on the context. In a second language context, like here, I don't think it's such a big drama, because they've got all the input around them. I think it's more difficult in a foreign language context, where input is limited, and when they're really dependent on one another. But I think probably learner-learner interaction has different kinds of goals and different benefits from whole-class teacher-led interaction, - it's interesting to think about what's the contribution of each to instructed language learning.

Sheila: How long have you been in the field?

Dr. Philp: I finished my PhD in 1998.

Sheila: In those 11 years have you noticed changes in the field or trends?

Dr. Philp: I think I was really lucky, the time that I started, because it was just at a time when people were starting to think about the processes behind language learning and looking at attention and working memory and all those kinds of things. I think over those ten years there's been a big push towards looking at processes of learning, and more recently thinking outside of just cognitive processes, looking at the social aspects of learning as well and trying to bring together many different pieces of the puzzle.