CHAPTER 43

# WORD PROSODY IN SECOND LANGUAGE ACQUISITION

ALLARD JONGMAN AND ANNIE TREMBLAY

## 43.1 Introduction

The learning of lexical prosody in adult second language (L2) learners differs greatly from that in children acquiring their native language (L1): not only are adults cognitively mature but they also approach the L2 learning task with an already established linguistic system, that of the L1. This system has an important influence on the learning of lexical prosody in an L2. The present chapter provides an overview of this influence by discussing L2 learners' use of prosodic information in word perception/recognition and their production of such information at the word level, focusing on the learning of lexical stress and lexical tone. The chapter first compares phonological and phonetic approaches to explaining L2 learners' ability to perceive and use stress in word recognition, and then compares phonological and statistical approaches to explaining L2 learners' production of lexical stress (§43.2). The chapter then discusses the learning of lexical tone by focusing on the contributions of lower-level acoustic-phonetic and higher-level linguistic information, the perceptual weighting of tonal cues, and the influence of contextual phonetic and prosodic information (§43.3). The efficacy of short-term auditory training is also discussed. The chapter ends with concluding remarks and future directions for research on L2 word prosody (§43.4).

## 43.2 Lexical stress

### 43.2.1 Second language word perception/recognition

One approach that has been adopted to explain the influence of the L1 on adults' perception of stress and use of stress in spoken-word recognition is the *phonological approach*. According to this approach, L2 learners' success at perceiving stress and using it in word

recognition is influenced by whether stress is represented phonologically as part of their L1 lexical representations. More precisely, L2 learners are more likely to use stress in word perception/recognition if stress is lexically contrastive in the L1 (i.e. if L1 words differ in their stress pattern) than if it is not lexically contrastive (e.g. Dupoux et al. 1997, 2001, 2008; C. Y. Lin et al. 2014; Peperkamp 2004; Peperkamp and Dupoux 2002; Peperkamp et al. 2010; Tremblay 2008, 2009). A number of studies have provided support for this approach.

Using AX perception and sequence recall tasks, Dupoux and colleagues showed that naive French listeners performed more poorly than native Spanish listeners when attempting to perceive stress in phonetically variable Spanish nonwords (i.e. nonwords differing in stress uttered by various Spanish speakers; e.g. Dupoux et al. 1997, 2001). Whereas Spanish words can differ in their stress pattern (e.g. Harris 1983), French words either are unstressed or have their final syllable 'stressed' in phrase-final position (e.g. Jun and Fougeron 2000, 2002).[1] French listeners' so-called stress deafness was attributed to stress not being lexically contrastive in French (Dupoux et al. 1997, 2001) (for similar results with speakers of Finnish, a language where stress is also not lexically contrastive, see Peperkamp and Dupoux 2002).

To provide a theoretical account of these findings, Peperkamp and Dupoux (2002) proposed the Stress Parameter Model (see also Peperkamp 2004). According to this model, listeners who, during the first two years of their life, receive exposure to a language in which stress is lexically contrastive (e.g. Spanish) set the Stress Parameter to encode (i.e. represent) stress phonologically in their lexical representations, whereas listeners who are exposed to a language where stress is not lexically contrastive (e.g. French, Finnish) do not. For listeners whose L1 does not have lexical stress, the model further predicts gradience in the degree of stress deafness as a function of whether the L1 prosodic system requires listeners to tease apart content words from function words. For example, in Hungarian, the first syllable of the first content word in a phrase is 'stressed' (Vago 1980), and in Polish the penultimate syllable of every content word is stressed (Comrie 1967). A lower degree of stress deafness is predicted for speakers of these languages than for speakers of languages where prosodic generalizations can be made independently of the lexical status of words (e.g. French, Finnish) (for such results, see Peperkamp and Dupoux 2002; Peperkamp et al. 2010) (for electrophysiological evidence that Polish listeners show different patterns of responses to different types of stress violations, see Domahs et al. 2012).

'Stress deafness' has been reported not only for naive French listeners but also for French-speaking L2 learners of Spanish. Dupoux et al. (2008) showed that French-speaking L2 learners of Spanish performed similarly to naive French listeners and significantly worse than native Spanish speakers when recalling the stress pattern of nonwords, and they performed significantly worse than native Spanish speakers when judging the lexical status of nonwords that differed from Spanish words only in their stress patterns, irrespective of their proficiency in Spanish. Even simultaneous French–Spanish bilinguals whose dominant language was French appeared 'deaf to stress' (Dupoux et al. 2010). On the basis of these findings, Dupoux et al. (2010) proposed that when the two languages learned from birth

[1] French does not have lexical stress; prominence is instead realized at the level of the phrase, with the last non-reduced syllable in the phrase receiving an intonational pitch accent and thus being perceived as more prominent than the preceding syllables (e.g. Jun and Fougeron 2000, 2002).

conflict in whether they have lexical stress, the Stress Parameter is set to encode stress only if the language with lexically contrastive stress is the dominant language.

The phonological approach to the study of stress in L2 word perception/recognition has received support from additional language pairings, including L1 French L2 English (Tremblay 2008, 2009) and L1 Korean L2 English and L1 Chinese L2 English (C. Y. Lin et al. 2014). In C. Y. Lin et al. (2014), for example, L2 learners of English whose L1 was Korean, a language where stress is not contrastive at the word level (Jun 2005b), had significantly more difficulty recalling nonwords that differed in stress compared to native English listeners and L2 learners of English whose L1 was (Standard) Mandarin, a tonal language where stress is lexically contrastive (Chao 1968; Duanmu 2007). These findings suggest that whether or not listeners can use stress in L2 word perception/recognition is strongly influenced by whether stress is lexically contrastive in the L1 (though not exclusively so; for details, see Peperkamp and Dupoux 2002).

The general predictions of the phonological approach are somewhat coarse, however, and the specific predictions of Peperkamp and Dupoux (2002)'s Stress Parameter Model have not been consistently supported (e.g. Rahmani et al. 2015). An approach that instead focuses on the specific cues that distinguish words from one another in the L1 and in the L2, henceforth referred to as the 'phonetic approach', may have more power in explaining L2 learners' use of stress in word perception/recognition. According to this cue-based, phonetic approach, adults' success at learning lexical stress in the L2 is also influenced by the degree to which the acoustic cues used to realize stress in the L2 (e.g. fundamental frequency (f0), duration, intensity, vowel quality) signal lexical contrasts in the L1 (e.g. Cooper et al. 2002; Zhang and Francis 2010; Ortega-Llebaria et al. 2013; Chrabaszcz et al. 2014; C. Y. Lin et al. 2014; Qin et al. 2017) (for a similar approach to the use of prosodic cues in L2 speech segmentation, see Tremblay et al. 2018). This approach has also received some empirical support.

Cooper et al. (2002), for example, showed that Dutch-speaking L2 learners of English were more accurate than native English listeners in a task where they selected the continuation of a stressed or unstressed word fragment they heard. Both English and Dutch have lexically contrastive stress, but unstressed syllables are more reduced in English than in Dutch (e.g. Sluijter and van Heuven 1996a). Since the word fragments used in the experiment all contained full vowels, listeners could not use vowel quality as a cue to lexical stress and thus had to rely on suprasegmental cues such as f0, duration, and intensity to determine whether the fragment was stressed and select the corresponding continuation for that fragment. Hence, the more accurate performance of the Dutch listeners (as compared to that of the English listeners) was attributed to their greater sensitivity to the suprasegmental cues to stress (see also van Heuven and de Jonge 2011). Cooper et al. (2002)'s results thus provide some support for a phonetic approach to the study of L2 learners' use of stress in word perception/recognition.

C. Y. Lin et al. (2014), discussed earlier, also obtained results that can be interpreted within a phonetic approach. In a lexical decision task, native English listeners, but not Korean- or Chinese-speaking L2 learners of English, were more likely to reject English nonwords that were incorrectly stressed if the incorrect stress placement affected the quality of the vowels in the word. The Korean listeners' results were attributed to the absence of vowel reduction in their L1. Although (Standard) Mandarin does have vowel reduction, reduced vowels cannot occur in word-initial syllables in Mandarin. The Chinese listeners' results

were attributed to the fact that many of the stimuli in C. Y. Lin et al. (2014)'s lexical decision task had a vowel quality change in the first syllable.

Stronger evidence for a phonetic approach to the use of stress in L2 word perception/recognition was provided by Zhang and Francis (2010). Using a word identification task with auditory stimuli in which the segmental and suprasegmental cues to lexical stress were orthogonally manipulated, the authors showed that although both native English listeners and (Mandarin) Chinese L2 learners of English relied more on vowel quality than on f0, duration, or intensity when recognizing English words that differed in stress, Chinese listeners' relative reliance on f0 was greater than that of English listeners. These results were attributed to the fact that Chinese has lexical tones and f0 is the primary cue to these tones (Howie 1976; Gandour 1983).

Additional support for a phonetic approach was provided by Ortega-Llebaria et al. (2013). Using a task in which listeners identified the stress pattern of a word in prenuclear position in a declarative sentence, the authors showed that English-speaking L2 learners of Spanish perceived syllables with an f0 rise as being stressed, unlike native Spanish listeners, who perceived the f0 rise as signalling stress post-tonically. Stressed syllables in prenuclear position in Spanish are associated with an f0 rise post-tonically (Hualde 2005; Prieto et al. 1995). L2 learners thus need to associate this f0 rise with the (stressed) syllable preceding it, something that Ortega-Llebaria et al. (2013)'s L2 learners did not appear to do. The L2 learners also made greater use of duration cues than did the native listeners. The duration ratio of stressed to unstressed syllables is larger in English than in Spanish (Delattre 1966), due in part to the occurrence of vowel reduction in English (Beckman and Edwards 1994) but not in Spanish. English listeners thus appeared to transfer the use of duration cues to the perception of stress in full vowels in Spanish. These results indicate that even when both the L1 and the L2 have lexically contrastive stress, L2 listeners must learn the acoustic cues to stress in the L2 in order to perceive stress accurately.

Chrabaszcz et al. (2014) provided further evidence that L2 learners' perception of stress is contingent on the cues that signal stress in the L1. In a stress perception task with nonwords, native English listeners and L2 learners of English who spoke Mandarin or Russian as their L1 were found to differ in their reliance on suprasegmental cues to stress: whereas both English and Mandarin listeners weighted f0 cues more heavily than duration and intensity cues, Russian listeners showed the opposite pattern of results. These results were attributed to the participants' L1, with f0 not being a reliable cue to stress in Russian, unlike in English and Mandarin.

Evidence in support of a cue-based approach was also provided by Rahmani et al. (2015). The authors reported that Dutch and Japanese listeners outperformed Persian, Indonesian, and French listeners when recalling sequences of nonwords that differed in stress. Japanese does not have lexical stress, but it has lexical pitch accents, with words differing in their tonal (i.e. pitch) patterns; in contrast, Persian does not have lexical stress or lexical pitch accents, and neither does Indonesian (for more details on the prosodic systems of each of these languages, see Rahmani et al. 2015). The authors interpreted their results as suggesting that listeners can encode stress in sequence recall tasks only if the L1 encodes prosodic markings at a lexical level. Although the authors' explanation is more phonological in nature, it yields the same predictions as a phonetic, cue-based approach, with Japanese listeners' use of pitch to differentiate L1 words enabling them to perceive and process L2 stress.

Qin et al. (2017) similarly showed that L2 learners' processing of lexical stress is modulated by which prosodic cues signal lexical contrasts in the L1. The authors investigated whether speakers of Standard Mandarin, a dialect with lexical stress (Chao 1968; Duanmu 2007), and speakers of Taiwan Mandarin, a dialect without lexical stress (Kubler 1985; Swihart 2003), differed in their ability to use f0 and duration cues when perceiving stress in English nonwords. In both varieties of Mandarin, f0 is the primary cue to lexical tones (Howie 1976; Gandour 1983), and in Standard Mandarin, duration is the primary cue to stress (T. Lin 1985). The results of a sequence recall task showed that, as predicted, L2 learners of English whose L1 was Standard Mandarin made greater use of duration cues when perceiving English stress than L2 learners of English whose L1 was Taiwan Mandarin, and both L2 groups made lesser use of these cues when compared to native English listeners. Crucially, when stress was realized with conflicting f0 and duration cues, both L2 groups relied more on f0 than on duration when perceiving English stress, whereas native English listeners relied equally on both types of cue. The greater reliance on f0 than on duration for Mandarin listeners was interpreted as their transferring the use of f0 from the perception of lexical tones in the L1 to the perception of stress in the L2. These findings thus provide further support for a cue-based, phonetic approach to the perception of L2 stress.

All in all, the existing research on the use of stress in L2 word perception/recognition suggests that listeners' success at perceiving stress in the L2 is predicted by both whether stress is lexically contrastive in the L1 and which prosodic cues signal lexical contrasts in the L1. Further research that focuses on the transfer of specific acoustic cues from the L1 to the L2 is needed in order to refine the predictions of the phonetic approach.

## 43.2.2 Second language word production

The influence of the L1 on adults' production of stress in the L2 has largely been studied from a *phonological perspective*. This research has typically focused on whether L2 learners from various L1 backgrounds stress the correct syllable in L2 words (e.g. Mairs 1989; Archibald 1992, 1993; Pater 1997b; Tremblay and Owens 2010). The general prediction from this approach is that L2 learners will be more successful at producing the correct lexical stress pattern if the generalizations underlying stress placement in the L1 (if any) are similar to those underlying stress placement in the L2.

Archibald (1992, 1993) analysed the stress systems of participants' L1 and L2 using the parameters of Metrical Theory, proposed by Dresher and Kaye (1990), and made predictions for the production of stress in L2 words using the L1 parameters. In a read-aloud task, Archibald (1992) found that when Polish-speaking L2 learners of English incorrectly stressed English words, they tended to stress the penultimate syllable. As mentioned earlier, Polish does not have lexically contrastive stress; in Polish, words are consistently stressed on the penultimate syllable. L2 learners' incorrect stress placement in English was attributed to Polish stress not being related to syllable weight (unlike in English, where syllables analysed as heavy should be stressed; Dresher and Kaye 1990) and to Polish words not ending in an extrametrical syllable (unlike in English, where the last syllable of nouns was analysed as extrametrical and, thus, invisible to stress; Dresher and Kaye 1990).

Using a similar task, Archibald (1993) showed that when Spanish-speaking L2 learners of English made stress placement errors in English words, they tended to stress either the penultimate syllable or the final syllable if the latter contained a diphthong or one or more coda consonants. Since the majority of the incorrectly stressed words ended with a derivational suffix, the author attributed these results to derivational affixes not being extrametrical in Spanish, unlike in English (for similar results, see Mairs 1989). Thus, in Archibald (1992, 1993), L2 learners' stress errors were attributed to the different generalizations underlying stress placement in the L1 and the L2.

Pater (1997b) adopted a similar approach with French Canadian L2 learners of English but instead elicited English nonwords. The nonwords were elicited in the subject position of a carrier sentence, and thus were interpreted as nouns. Pater (1997b) found that these L2 learners typically stressed the first syllable of trisyllabic nouns and, unlike native English speakers, showed little sensitivity to syllable structure in their production of lexical stress.[2] Importantly, the L2 learners rarely stressed the last syllable of the trisyllabic nonwords, which is the pattern of responses that Pater (1997b) predicted based on his analysis of Canadian French using Dresher and Kaye (1990)'s parameters (with French being analysed as having an iambic, quantity-insensitive foot), suggesting a lack of L1 transfer.

Tremblay and Owens (2010), who used a similar task, also reported a tendency for French Canadian L2 learners of English to stress the initial syllable of disyllabic and trisyllabic nonce nouns, independently of syllable structure. The results of Pater (1997b) and Tremblay and Owens (2010), unlike those of Archibald (1992, 1993), suggest that L2 learners do not necessarily show clear evidence of L1 influence in their production of L2 lexical stress. Tremblay and Owens (2010) attributed the L2 learners' production of initial stress to the statistical frequency with which nouns are stressed on the initial syllable in English (e.g. Cutler and Carter 1987; Clopper 2002), leading them to overgeneralize this pattern to contexts where stress should not be word-initial (e.g. in trisyllabic nouns that contain a heavy penultimate syllable).

L2 learners' production of lexical stress has also been investigated from a *statistical perspective*. This approach has focused on whether L2 learners can learn the statistical regularities of stress patterns in the L2, independently of the L1 (e.g. Davis and Kelly 1997; Guion 2005; Guion et al. 2004). The researchers who have conducted studies under this approach have expressed concerns about the psychological reality of the stress rules proposed to explain stress placement, at least in English (for discussion, see Guion et al. 2004).

Guion et al. (2004) examined the production of English nonwords by Spanish speakers who had acquired English at an early age (mean of 3.7 years) or at a later age (mean of 21.5 years).[3] Statistically, in English, disyllabic words are more likely to be stressed initially if they are nouns than if they are verbs (Sereno 1986; Kelly and Bock 1988), and syllables with diphthongs are more likely to be stressed than syllables with lax vowels (Guion et al. 2003). Guion et al. (2004) showed that the early L2 learners (and native English speakers) were more likely to stress the first syllable of disyllabic nonce words when they were elicited as nouns than when they were elicited as verbs. Furthermore, the early L2 learners

[2] Pater (1997b) also examined L2 learners' production of secondary stress in quadrisyllabic words. We do not discuss the results for these words due to space limitations.

[3] Guion et al. (2004) also elicited judgments of stress patterns in the same nonwords. We do not report these results due to space limitations.

(and native speakers) were more likely to stress syllables that contained a diphthong than syllables that contained a lax vowel, and to stress syllables that contained a complex coda than syllables that contained a simple coda. The late L2 learners, on the other hand, showed an effect of lexical class only for words that contained a lax vowel in the first syllable and a simple coda in the second syllable, and an effect of vowel only for words that had a diphthong in the second syllable. They were also more likely to stress the first syllable of all words compared to the early L2 learners and native speakers, suggesting a similar overgeneralization of the word-initial stress pattern as observed in Pater (1997b) and Tremblay and Owens (2010). The authors interpreted their results as suggesting that the L2 learners' age of acquisition had an important effect on their ability to extract the statistical regularities of stress patterns from the input, especially regularities that relate to syllable structure.

In an attempt to determine whether Korean speakers could learn the statistical regularities of stress patterns in English, Guion (2005) conducted a replication of Guion et al. (2004)'s study but with early and late Korean L2 learners of English. Her production results were similar to those of Guion et al. (2004), with age of acquisition having an effect on L2 learners' ability to extract the stress regularities that relate to lexical class and syllable structure from the input. However, since the two studies were reported in different papers, it is unclear whether the L1 significantly affected L2 learners' production of stress.

In summary, the existing research on the production of lexical stress in the L2 suggests that L2 learners can but do not necessarily transfer the generalizations underlying stress placement from the L1 to the L2. Furthermore, L2 learners appear to be able to learn the statistical regularities that relate to stress placement in the L2, though not at a native-like level, with early L2 learners outperforming late L2 learners.

## 43.3 Lexical tone

L1 effects have been found not only on the perception/recognition and production of lexical stress but also on the perception/recognition and production of lexical tones. It is by now well established that there is a difference in the way native speakers of a tone language perceive tonal distinctions as compared to native speakers of a non-tonal language. Most generally, speakers of a tone language can discriminate tones more accurately and quickly than speakers of a non-tonal language (see Lee et al. 1996; Wayland and Guion 2003; Bent et al. 2006).

Furthermore, speakers of a tone language and speakers of a non-tonal language are differentially sensitive to individual pitch cues. In a tone language, tone (or its acoustic correlate, f0, or its perceptual correlate, pitch) serves to distinguish word meaning. In a non-tonal language, on the other hand, while pitch may provide grammatical or intonational information, it does not distinguish word meaning. Speakers of a tone language are more sensitive to changes in pitch direction and slope, which are crucial to tonal identification and thus linguistically relevant, whereas speakers of a non-tonal language attend more to general phonetic properties such as pitch height (average pitch) and duration, which are arguably linguistically less relevant in the languages under investigation (Gandour and Harshman 1978; Gandour 1983; Chandrasekaran et al. 2007b; Jongman et al. 2017).

Additionally, studies of categorical perception of tones show that tonal distinctions that involve pitch movement are perceived in a categorical manner by native speakers of tone languages (e.g. Hallé et al. 2004; Xu et al. 2006a; Peng et al. 2010). Listeners of non-tone languages, on the other hand, are much more sensitive to within-category differences as compared to native speakers of tone languages. This less categorical nature of tone perception has been reported for listeners from a variety of non-tonal-language backgrounds, including Dutch (Leather 1987), English (Xu et al. 2006a), French (Hallé et al. 2004; see also DiCanio 2012b for French participants listening to Trique), and German (Peng et al. 2010). It has been suggested that speakers of non-tonal languages process the stimuli in an acoustic or psychophysical mode, such that they are sensitive to small acoustic differences as long as they exceed the difference limen (e.g. Hallé et al. 2004); in contrast, speakers of tone languages process the stimuli in a linguistic mode, ignoring small acoustic differences between members of the same category so that they can assign them to one of two tonal categories.

The above-reviewed differences in tonal perception between speakers of a tone language and speakers of a non-tone language entail challenges in tonal acquisition in an L2. In what follows, we will discuss a number of these challenges as well as the extent to which training can mitigate them.

### 43.3.1 Second language perception/recognition of lexical tone

L2 tonal acquisition has been shown to be affected by several factors, including training, familiarity with lexical tone, language proficiency, and tonal transfer. To begin with, research to empirically assess learner performance employs the well-established high-variability phonetic training paradigm aimed at assisting learners to establish L2 phonetic categories by exposing them to a great variety of exemplars of a category (Logan et al. 1991). In one of the first tone training studies, Y. Wang et al. (1999) demonstrated that a brief training regimen (eight sessions over the course of two weeks) significantly improved English learners' identification of Mandarin tone. During training, learners (students in their first semester of Mandarin instruction) were exposed to a variety of talkers and phonetic contexts; tones were trained pairwise and learners received feedback. The results showed a substantial gain (21%) in tone identification accuracy after training, as compared to no gain in a control group that received no training. All four tones were identified more accurately after training. In addition, the training benefit extended to both words and speakers not encountered during training, and was still present six months after training.

Training has also been extended to native speakers of a tone language learning a foreign tone system. While training in the laboratory clearly improves learners' perception of tone, it is important to establish whether this improvement reflects more native-like processing. Francis et al. (2008) provided Mandarin Chinese (tonal) and English (non-tonal) participants with 10 hours of training on Cantonese tones. Overall, training improved the performance of both groups to the same extent. Multi-dimensional scaling analysis showed that the two primary dimensions, pitch height and pitch slope, accounted for a greater proportion of the variance after training compared to before training. That is, both groups

became more Cantonese-like in their weighting of these two cues. However, language-specific differences remained, with non-tonal English participants giving more weight to height over direction while Mandarin participants assigned about equal weight to both cues.

Wayland and Guion (2003) addressed the role of language proficiency by comparing the discrimination of the mid and low tones in Thai by native speakers and by English speakers with and without Thai language experience. The experienced speakers had studied Thai for 2.5 years and had lived in Thailand for four years on average. The performance by the native speakers of Thai was best; the English speakers without Thai experience did the worst, and the L2 learners of Thai fell in between. Similarly, Guion and Pederson (2007) showed that, when discriminating tone, advanced American learners of Mandarin were more similar to native Mandarin speakers than Americans without any experience with a tone language: while the last group only used average pitch in their discrimination of synthesized stylistic tones, both the advanced learners and the native speakers of Mandarin used pitch slope in addition to average pitch.

Qin and Jongman (2016) directly evaluated the role of tone transfer by comparing discrimination of Cantonese level and contour tones by native speakers of Mandarin, native speakers of English, and English learners of Mandarin. They found that both the native Mandarin speakers and the English learners of Mandarin were better at discriminating the contour–level tone pairs than the level–level tones. This suggests that experience with Mandarin increased L2 learners' sensitivity to f0 direction in the perception of Cantonese tones. Moreover, the L2 learners, as well as the monolingual English speakers, were better than the native Mandarin speakers at discriminating the level–level tone pairs, suggesting that the English L1 experience still influenced how L2 learners of Mandarin perceived Cantonese tones.

Overall, previous lexical tone experience in a tone system, be it either as an L1 or an L2, transfers to the perception of tones in a different tone system. In addition, tone training improves tone categorization even for participants with little or no exposure to a tone language. However, differences in the weighting of cues to tone may remain between learners and native speakers of a tone language, and between learners from different non-tonal L1 backgrounds (e.g. Braun et al. 2014). Brain imaging studies can provide an additional way of assessing the extent to which the improvement observed after training reflects more native-like processing (e.g. Kaan et al. 2008).

## 43.3.2 Second language production of lexical tone

The production of tone has received much less attention than the perception of tone. While much research remains to be done in this area, there are several studies that report on the difficulties that L2 learners of a tone language encounter. For example, Shen (1989b) found that English learners of Mandarin had difficulty with the production of all tones but especially with Tone 4. Tone 4 errors may be ascribed to the fact that this falling tone is less prosodically marked for speakers of English, who may therefore use it more frequently to substitute for other tones. Miracle (1989) conducted an acoustic analysis of productions made by second-year American learners of Mandarin. She used the pitch contours of the learners as well as three native speakers to assign the productions to one of the four tonal categories. Miracle (1989) reported an error rate of 43%. These errors were evenly divided

between register (too low or high) and contour errors and across all four tones. While Y. Wang et al. (2003a) observed a similar overall error rate as Miracle (1989), the distribution of errors was quite different, with accuracy scores for Tones 1, 2, and 4 around 67% but only around 20% for Tone 3. It should be noted that Tone 3 is also acquired last by native speakers of Mandarin (e.g. C. N. Li and Thompson 1977).

Y. Wang et al. (2003a) also directly examined the tone production of their participants in a perceptual training study (Y. Wang et al. 1999). The training group read a wordlist before and after training while the control group read this wordlist twice separated by two weeks. Pre- and post-training productions were evaluated in two ways: by native Mandarin listeners in an identification task and by acoustic measurements. Results from the perceptual evaluation indicated that a significantly greater number of tones produced after training were perceived as intended as compared to those produced before training. The acoustic analysis consisted of a detailed comparison of a number of parameters including onset, offset, maximum, and minimum pitch value between native Mandarin and learner productions. The acoustic results were consistent with the native-speaker judgements as they showed that the pitch contours of the learners' productions were closer to those of the native speakers after training as compared to before training. This finding is particularly interesting when it is considered that the participants in this study were only trained on the perception of tone and did not receive any production training. Nevertheless, the benefits from the perception training seemed to carry over to production.

## 43.4 Conclusions and future directions

The above research shows important L1 effects on how adult listeners perceive and process lexical stress and lexical tone in L2. For lexical stress, listeners' success at encoding stress in the L2 is predicted by both whether stress is lexically contrastive in the L1 and which prosodic cues signal lexical identity in the L1. For L2 learners whose L1 does not have lexically contrastive stress, proficiency in the L2 does not seem to be a strong predictor of listeners' ability to perceive and use lexical stress in word recognition. A better predictor is instead whether a prosodic cue signals lexical identity in the L1, with L2 learners relying on this cue to perceive and use stress in L2 word recognition even if the L1 does not have lexically contrastive stress. For lexical tone, speakers of tone languages and speakers of non-tonal languages differ in the weighting of tonal features, with speakers of tone languages being more sensitive to changes in pitch direction while speakers of non-tonal languages attend more to pitch height. Native knowledge of a tone language or acquired proficiency in a tone language can help the processing of tone in an L2. However, the best predictor of performance in a non-native tone language may be the extent to which the acoustic features that characterize the tones in the L1 correspond to those used in the tones of the L2. Although the research cited in this chapter also reveals L1 effects on the production of lexical stress and lexical tones, L2 learners do not necessarily show L1 effects in their production of lexical stress and may instead be sensitive to the statistical regularities that relate to stress placement in the L2.

Laboratory training studies using the high-variability training paradigm have been shown to substantially improve learners' perception and production of tone. Such research should be extended to the perception and production of lexical stress by L2 learners whose L1 does not have lexically contrastive stress (for an example of a study with English L2 learners of Spanish, see Romanelli et al. 2015). A key challenge is to establish ways to effect a shift in cue weighting to attain more native-like performance (e.g. Lim and Holt 2011). For lexical tones, while much of the training research has focused on monosyllabic words, future research should expand to polysyllabic words in which tonal coarticulation creates a much greater degree of acoustic variability. For example, it has been shown that it is more difficult to learn tones in disyllabic than monosyllabic words and that the learnability of a given tone is correlated with its contextual tonal variability (Chang and Bowles 2015). Moreover, preliminary results suggest that training on monosyllabic words does not result in improved tone perception in bisyllabic words, while training on bisyllabic words does transfer to monosyllabic words (Y. Li 2016). Preliminary findings also show that tone training at phrasal and sentential levels results in improvements in tone perception in these larger linguistic contexts (X. Wang 2012).

The roles of individual differences and aptitude need to be investigated in more detail. Research indicates that individual differences in cue weighting may be able to predict which participants benefit most from training (Chandrasekaran et al. 2010), and perceptual aptitude may determine which type of training is preferable, with high-variability training benefiting high-aptitude perceivers while low-aptitude perceivers preferred low-variability training (e.g. Perrachione et al. 2011; Sadakata and McQueen 2014).

Finally, there are additional promising areas that could not be included in this review because of space constraints. One such area concerns L1–L2 interactions in tone perception in bilinguals. For example, experience with a tonal language has been shown to affect the processing of English intonation by Chinese learners of English (Ortega-Llebaria et al. 2015). Another area requiring more research is the extent to which the communicative goal (e.g. focusing on tone or intonation) makes listeners shift their cue weighting (X. Li et al. 2008).