3

HERITAGE VERSUS SECOND LANGUAGE PHONOLOGY

Yasaman Rafat

3.1 Introduction

This chapter will focus on discussing some of the different ways in which Persian heritage phonetics and phonology have been analyzed, and will draw parallels with some of the previously reported patterns in second language (L2) speech learning. Although there is a rich body of literature on L2 speech learning (e.g., Best and Tyler 2007; Brown 1998; Colantoni and Steele 2007; Flege 1995), less is known about heritage speech. The need to expand on the scarce research on heritage speech also applies to Persian heritage and L2 speech. Chapter 4 in this volume is an overview of the research previously conducted on Persian heritage linguistics, with a focus on the domains of phonology, morphology and syntax. It compares the linguistic competences of Persian heritage versus the second language learners of Persian.

The rest of this chapter is organized as follows. Section 2 offers various definitions of heritage speakers and calls for a need for a definition that depicts heritage speech as a more transient phenomenon. Section 3 provides a more holistic approach to viewing both L2 speech learning and heritage speech as a multimodal event, highlights the role of orthography in L2 speech learning, and points out some of the differences in the different modalities with respect to Persian and English. Heritage speech is analyzed from a more diachronic and sociophonetic lens in Section 4, where Persian heritage speech is considered in the context of language change across generations. This section also draws parallels between heritage and L2 speech. Section 5 discusses heritage speech from a developmental point of view and discusses Persian-English-speaking children. Section 6 concludes the chapter and provides future research directions. This chapter reviews evidence on characteristics of heritage speech and L2 speech in order to contextualize the existing studies on two infrequently studied topics: Persian heritage speech and L2 speech. For further research on the Persian heritage learners versus second language learners of Persian, read Chapter 4 in this volume, where there is an elaborate comparison made between these two groups of learners in the domains of phonology, morphology and syntax.

3.2 Heritage speakers

There is currently no consensus on what constitutes a heritage speaker. The definition of heritage speaker and heritage language (HL) started to be developed during the ’70s in Canada, but
Heritage versus second language phonology

It gained the attention it deserves only twenty years later. Regardless of the definition chosen, it is important to remember that typically heritage speakers have a first and a second language (order of acquisition), a primary and a secondary language (functional dimension), a majority and a minority language (sociopolitical dimension; Montrul 2012).

Draper and Hicks (2000) state that a heritage speaker is someone who has been exposed to a non-English language outside the formal educational schooling setting. They also included speakers who grow up speaking a different language at home and speakers with a strong in-depth exposure to another language. This definition is a very broad description of heritage speakers.

While Campbell and Peyton’s (1998) definition of a heritage speaker as an individual who speaks a first non-English language at home or who is born in a different country is widely accepted, they seem to have overlooked an important connection made later by Valdés (2001). That is the individual’s personal connection with the language. Valdés (2001) suggested that heritage speakers are individuals who have a historical and/or personal connection to a language that is not normally taught at school, e.g., indigenous or minority. She added that heritage speakers are individuals who grow up speaking a non-English language at home, with all levels of proficiency in the heritage language and all levels of bilingualism. Valdés’ (2001) main goal however is strictly pedagogical.

Recently, Montrul (2012) proposed that heritage speakers are early bilinguals, because of the exposure they have had to both the heritage and the majority language. As bilinguals, heritage speakers can be simultaneous when they grow up speaking both languages, or sequential when they learn the second language after the age of five or six. She suggested that regardless of when a heritage speaker starts learning both languages, by the time they enter adulthood, the heritage language is weaker.

Recently a new debate has ensued about whether heritage speakers should be considered native speakers (Rothman 2006, 2009; Rothman and Treffers-Daller 2014; Kupisch and Rothman 2018). Rothman’s definition for heritage speakers starts where Montrul’s ended: heritage speakers are bilinguals, but not all bilinguals are heritage speakers. According to his studies, heritage speakers are on the continuum between monolingual and bilinguals, as they have been naturally exposed to a multilingual environment since childhood, if not even birth, and have competence in both languages. Rothman (2009) suggested the idea that the essential condition for heritage speakers is the naturalistic environment in which they need to learn the language. Rothman (2006) has suggested that heritage speakers can present different levels of proficiency, depending on personal and social factors. More recently, Kupisch and Rothman (2018) divided heritage speakers into two categories: early bilinguals, who have been exposed to the language in a naturalistic environment outside of the school setting and have strong personal connection to the language such as family, culture and intrinsic motivation, and those who learned the language as adults and therefore lack the naturalistic environment. In their opinion, the former and the latter can also be considered native speakers of the home language (Rothman and Treffers-Daller 2014; Kupisch and Rothman 2018), regardless of the level of proficiency or the hypothesized incomplete acquisition that Valdés (2000) and Benmamoun (2013) have proposed.

Although there is not a single definition on what constitutes a heritage speaker, these definitions show that heritage speakers are at some point in their lives exposed to two different languages although their degree of proficiency of each language, the order of acquisition of the two languages, and their affinity with the two cultures might vary. However, these definitions present a static rather than a more dynamic picture of what a heritage speaker might be. For example, all parents of heritage speakers know that their children’s proficiency in both
their heritage (e.g., Persian) and the majority language (e.g., English) may vary at a given time. This is highly dependent on the amount and kind of exposure that heritage speakers may have to both languages. To present an extreme situation, child heritage speakers, who barely speak their heritage language, if and when they go back to their country of origin for a summer visit, may become fluent in their heritage language and forget their other language. However, they might start becoming more dominant in English after a couple of weeks of exposure to it again in school. Moreover, Choi, Boersma, and Cutler (2017) found that adoptees, who have been exposed to a home language in infancy for a very short period of time (e.g., 3–5 months), exhibit an advantage in the ability to relearn the language as adults in comparison with controls. The authors concluded that early exposure to spoken language, even in the first half-year of life, may leave traces that can facilitate later relearning. Although adoptees are not typically considered heritage speakers, a more dynamic definition of a heritage speaker and a model of heritage speech, which contextualizes heritage speakers and heritage speech learning and change across the lifespan, are called for.

In the past decade, the body of evidence has been growing on heritage speakers, although more has been done on morpho-syntax (e.g., Montrul 2005, 2006, 2007a, 2007b) and less on phonetics or phonology (e.g., Au et al. 2002; Rafat, Mohaghegh, and Stevenson 2017). However, as Kupisch and Rothman (2018) wisely pointed out, definitions do matter and there is an urge to continue examining heritage languages, the same way scholars have focused on L2 learners. This chapter considers heritage speakers to be speakers who, very early on in life or as children, have had some exposure to a home language that is different from the dominant language in the societ(ies) that they grew up in, and have some passive or active knowledge of their home language. Similarly to L2 learners, heritage speakers may differ in their proficiency level and exhibit individual differences. Furthermore, their proficiencies in their heritage and dominant languages may differ across their lifespan. The degree of their proficiency in their heritage language may be driven by both linguistic and extra-linguistic factors. The latter includes the amount of input, the degree to which they may identify with their heritage culture at any given time in their life, and other social factors such as age, gender, education, socio-economic background and the density of their network.

3.3 L2 and heritage speech learning as a multimodal event: the role of orthography

Currently, there are a number of models of the acquisition of L2 phonetic and phonological acquisition (e.g., Best and Tyler 2007; Brown 1998; Colantoni and Steele 2007; Flege 1995). However, no specific models explain heritage speech learning. Instead, the Speech Learning Model (SLM; Flege 1995) has been applied to heritage speech learning. Flege’s SLM is a perception-based model that predicts that the degree of acoustic difference between the L1 and L2 sounds will determine whether the L2 sound will be mapped on to the L1 sound. In other words, the more dissimilar the sound or the larger the acoustic-phonetic distance between the L1 and the L2, the higher the possibility that this sound will be acquired. On the other hand, the smaller the acoustic-phonetic distance between the L1 and the L2 sounds, the more likely that equivalence classification will take place and the L2 will not be perceived as distinct. For further discussions on the Speech Learning Model, read Chapter 2 in this volume, which employs the theoretical framework of the SLM to test the Persian L2 data against different approaches.

Another well-known model is Best and Tyler’s (2007) PAM-L2, a revised version of Best’s (1995) Perceptual Assimilation Model (PAM). PAM-L2 is based on the premise that sounds are perceived in terms of articulatory gestures (e.g., Browman and Goldstein 1989, 1990,
Heritage versus second language phonology

1992, 1995). Moreover, the patterns of assimilation to L1 categories determine the accuracy of the discrimination of target language (TL) contrasts and subsequent category formation. Very good to excellent discrimination is predicted for Two Category assimilation, in which two TL phones are perceived as acceptable exemplars of two different L1 phones; poor discrimination is predicted when two TL sounds are perceived as equally good or poor exemplars of the same L1 phoneme; and Single Category assimilation and intermediate discrimination is predicted when two TL sounds differ in the extent to which they are good exemplars in relation to a single L1 phoneme. Of all the formal models of (perceptual) speech learning, this model is the only one that briefly mentions the effect of orthography on the categorization of TL sounds.

Whereas both the SLM and the other models of L2 speech learning have been mostly perception-based, Colantoni and Steele (2007) consider both perception and production. They argue that because previous L2 speech learning models do not consider the degree of difficulty involved in the simultaneous mastery of multiple phonetic parameters across prosodic positions, they do not account for the full range of variability nor for the developmental sequences attested.

Despite our understanding that many aspects of speech processing are multimodal, including speech perception (e.g., Sumby and Pollack 1954; McGurk and MacDonald 1976; Massaro 1987; Soto-Faraco, Navarra, and Alsius 2004), speech learning (Vigliocco, Perniss, and Vinson 2014), and second language (L2) speech learning (Hardison 1999; Ortega-Llebaria, Faulkner, and Hazan 2001; Erdener and Burnham 2005, 2013), the earlier-mentioned models of L2 speech learning are for the most part based on auditory-input only. One of the most salient examples of the multimodal nature of speech is the McGurk effect (e.g., McGurk and MacDonald 1976; Welch and Warren 1980; Sekiyama and Tohkura 1991; Munhall et al. 1996; Sekiyama 1997). The McGurk effect is elicited by the synchronous or simultaneous presentation of incongruent auditory (e.g., /ba/) and facial/visual cues (e.g., /ga/). The listener often integrates the auditory and visual information leading to either (a) a combination percept, such as /bga/ (McGurk and MacDonald 1976; Green and Norrix 1997) suggesting a strong influence of vision; or (b) a fused percept, such as /da/, where the syllable perceived is not contained in either the auditory or the visual information (e.g., Green and Kuhl 1989; Green et al. 1991; MacDonald and McGurk 1978; Manuel et al. 1983; Massaro 1987; Sekiyama and Tohkura 1991; Summerfield and McGrath 1984; Stevenson et al. 2014). Both combination and fused perceptions are different from the individual original sounds presented separately in each modality, not only in that they are different phonemes, but that they result from an interaction between sensory modalities.

There is also an abundance of research that has provided evidence for the orthographic (writing) channel interacting with auditory input in L1 speech processing (Dijkstra, Roelofs, and Fieuws 1995; Jakimik, Cole, and Rudnicky 1985; Montant et al. 2011; Seidenberg and McClelland 1989; Seidenberg and Tanenhaus 1979; Van Orden and Goldinger 1994; Ranbom and Connine 2011; Taft 2006; Treiman and Cassar 1997; Ziegler and Ferrand 1998; Ziegler and Muneaux 2007, among others). For instance, Seidenberg and Tanenhaus (1979) claimed that orthographic knowledge can affect spoken word processing. They conducted a rhyme judgment task and found that participants’ responses were faster in a rhyme when the pairs of words shared spellings (e.g., <toast> – <roast> vs. <toast> – <ghost>). Similarly, Jakimik and colleagues (1985) conducted priming tasks and found that the participants responded faster to the auditory target with the prior presentation of a phonologically similar prime that overlapped with the spelling of the target (e.g., message – mess) than to a prime with non-overlapping spellings (e.g., <definite> – <deaf>). Orthography has also been shown to affect underlying representations (Ranbom and Connine 2007, 2011). Ranbom and Connine (2007),
for example, provided some evidence that orthographic information affects mental representations of speech, specifically, the representation of lexically stored allophonic representations. They conducted a corpus analysis and found that the nasal flap realization in the /nt/ cluster of the word *gentle* is dominant in spoken American English, even though the production frequency of the nasal flap may vary within individual words. They then conducted a lexical decision task and showed that the highly frequent nasal flap was identified more quickly and accurately than the less frequent flap, but, crucially, [nt] productions resulted in faster and more accurate lexical decisions compared with the nasal flaps. The results of the lexical decision task demonstrated that orthographic information influences spoken word processing.

There is also some evidence to suggest that orthography may exert an influence on spoken word production (Bentur 1978; Ravid and Shlesinger 2001; Temkin Martinez and Müllner 2015; Han and Choi 2016). For example, Han and Choi (2016) investigated the role of orthography in production and storage of spoken words by Korean speakers. The participants learned novel Korean words with different variants of /h/ including [ɦ] and [ø]. They were provided with the same auditory stimuli but different exposures to orthography. There were two orthographic groups and an auditory-only group. One orthographic group was presented the letter for [ɦ] (<orraine>) and the other with the letter for [ø] (<ror>). The auditory group was presented with auditory input only. In picture-naming tasks, the participants presented with <ror> produced fewer words with [ɦ] than those presented with <orraine>. In a spelling recall task, the participants who were not exposed to spelling displayed various types of spellings for variants, but after exposure to spelling, they began to produce spellings as provided in the task. These results were attributed to orthographic input influencing production because of its potential to restructure phonological representations. For further research on priming tasks and processing in Persian, read Chapter 6, which delves into the processing of idiomatic expressions in L1 and L2 Persian speakers and discusses the role of morphology and orthography on the processing of such expressions.

The body of literature has also been expanding with respect to how orthography may interact with acoustic-phonetic input in L2 speech learning. Notably most research has focused on the Roman alphabet. When the learner’s L1 and L2 share the same alphabet, learners are faced with two main challenges. First, they have to learn that the L1 and the L2 mappings might be different. The correspondence of one grapheme to different L1 and the L2 sounds often leads to L1-based transfer (Rafat 2011, 2015, 2016). Second, learners may need to learn one-to-many mappings (e.g., <x> in Spanish may map on to [ks] in the word *taxi* but to [x] in *México* and to [gz] in *examen*) or many-to-one mappings in the L2 (e.g., <x>, <gi>, <ge> and <j> correspond to [x] or its other variants in Spanish). The orthographic depth hypothesis (ODH; Katz and Frost 1992) postulates that speakers of languages with shallow/regular/transparent orthographies tend to be more affected by orthographic input than speakers of languages with deep/irregular/opaque orthographies. However, we still do not know the extent to which L1 orthographic depth may modulate orthographic effects in L2 learners whose L1 orthographic system is alphabetic (e.g., Erdener and Burnham 2005; Rafat 2015; Escudero 2015).

English and Spanish both have a Roman alphabetic system. Whereas the English orthographic system is characterized by one-to-many grapheme-to-phoneme correspondences and is therefore considered an irregular/deep orthography, the Spanish orthographic system is mainly characterized by one-to-one mappings. However, L2 speech learning by English-speaking learners of Spanish exhibits orthographic effects due to the differences between English and Spanish grapheme-to-phoneme correspondences (e.g., Rafat 2011, 2015, 2016).

Orthographic effects have also been shown in several L2 perception and production studies. These studies have demonstrated that orthographic input may interact with auditory input and
may promote (e.g., Erdener and Burnham 2005; Steele 2005; Showalter and Hayes-Harb 2013; Bassetti, Escudero, and Hayes-Harb 2015; Rafat 2015), or hinder (e.g., Bassetti 2007; Erdener and Burnham 2005; Hayes-Harb, Nicol, and Baker 2010; Young-Scholten 2000; Young-Scholten, Akita, and Cross 1999; Bassetti, Escudero, and Hayes-Harb 2015; Nimz 2016; Rafat 2011, 2015, 2016; Bassetti 2017; Shea 2017) the target-like production or correct perception of the target L2 sounds, or have no effect (Escudero 2015; Showalter and Hayes-Harb 2013).

Young-Scholten, Akita and Cross (1999) conducted one of the first studies on the effect of orthography on L2 speech production. They examined the effect of exposure to orthographic input on the production of Polish clusters by adult English-speaking L2 learners. They found that exposure to orthography results in less omission and more epenthesis as in older learners.

Hayes-Harb, Nicol, and Baker (2010) examined the interfering effect of orthographic input in novel word learning by English-speaking participants when the grapheme-to-phonemes in the target language do not match. Participants were assigned to three different conditions at training: auditory-only, congruent, and congruent/incongruent orthography. The incongruent stimuli consisted of items spelled with a “wrong” letter (e.g., < faza> – [faʃə]) and items with an extra letter (e.g., < kamand> – [kaməd]). During testing, participants were shown an image and heard a word, and then asked whether the auditory word was the correct word for the image. The results yielded a significant effect of training condition on performance on the wrong-letter items, in which participants had a lower rate of accuracy in the incongruent/congruent orthography condition.

Mathieu (2016) also examined the effect of orthography at the onset of the acquisition of an L2. He reported the effects of three L2 scripts on the early acquisition of an Arabic consonantal contrast word-initially (e.g., /ħal/ – /χal/), showing that foreign written input can inhibit learners’ ability to encode an L2 phonological contrast. He tested monolingual native speakers of English with no prior knowledge of Arabic. Participants took part in a word-learning experiment and were assigned to one of four learning conditions: no orthography, Arabic script, Cyrillic script or Roman/Cyrillic blended script. The results showed that the degree of script unfamiliarity does not in itself seem to significantly affect the successful acquisition of the phonological contrast tested. However, the presence of certain foreign scripts in the phonological acquisition yielded significantly different learning outcomes in comparison to having no orthographic representation available. Specifically, the Arabic script exerted an inhibitory effect on L2 phonological acquisition, while the Cyrillic and Roman/Cyrillic blended scripts exercised different inhibitory effects based on whether grapheme-phoneme correspondences triggered L1-based phonological transfer. Mathieu (2016) proposed that L2 speech learning may be multimodal and subject to instantaneous and automatic processing similar to the McGurk effect. Specifically, the processing of the visual input at the early stages of acquisition may prompt the auditory system to strengthen activation of the L1 phonological categories, in turn hindering the accurate perception of the novel L2 sounds.

The effect of orthographic input has also been shown in more advanced learners. Bassetti (2007) investigated the effect of orthographic inconsistency within the L2 on the production of triphthongs by Italian-speaking learners of Mandarin studying at a university in Italy. Participants used the alphabetic pinyin writing system and on average had studied Mandarin for 33 months. Although the participants had not been exposed to pinyin orthography during the character-reading task, the results yielded a 100% target-like realization of the vowel /o/ in the triphthong /iou/ when it was written with three graphemes as in <you>. However, erroneous productions were attested when the triphthongs were spelled with only two graphemes in pinyin (e.g., <iu> for /iou/). The author explained the results by proposing that pinyin generally is a transparent orthographic system, and the learners had overgeneralized this aspect of pinyin.
There is considerable evidence that when the L1 and L2 grapheme-to-phoneme relationships are incongruent, exposure to orthographic input may result in L1-based phonetic or phonological transfer (e.g., Young-Scholten 2002; Rafat 2011, 2015, 2016; Bassetti 2017). An example can be observed in the production of digraphs (two graphemes such as <tt> in <kitty>) by highly proficient Italian-speaking learners of English. These learners produce digraphs as long/geminate consonants with a significant difference between their CC vs. C productions in English (Bassetti 2017). The authors attributed this to transfer of L1 phonological rules, in this case gemination.

Although several studies have provided evidence of transfer effects, this effect is not categorical and may be modulated by various factors. For example, transfer in English-speaking learners’ devoicing of syllable-final consonants is argued to be modulated by the amount of exposure to orthographic input in German (Young-Scholten 2000). In German, obstruents are devoiced in syllable-final position, although this is not cued in the orthography. For example, the word /bʊnd/ “federation” is written as <bund> but is realized as [bʊnt] creating a homophone with [bont] “coloured,” which is written as <bunt>. When students learn German as an L2, greater exposure to written text is related to a reduction in the acquisition of this obstruent final devoicing rule (Young-Scholten 2000). Evidence has also been found for orthography-induced transfer in the word initial /z/ production of the same group of learners, where they produced <s>, which corresponds to /z/ in German as [s] (e.g., [siː] for <sie> “she” /ziː/) (Young-Scholten and Langer 2015). An acoustic analysis of the results also revealed evidence for some partially voiced versions of /z/, which the authors suggested is a reflection of the variability in the auditory input that the learners had been exposed to.

Other factors such as type of grapheme-to-sound correspondence, position in the word, and condition of training and testing have been reported to control the rate of orthography-induced transfer in naïve English-speaking learners of Spanish (e.g., Rafat 2011, 2016). Exposure to orthographic input at the time of learning yielded a significantly higher rate of transfer compared to when orthographic input was presented at production or testing only. Additionally, different grapheme-to-sound correspondences resulted in significantly different rates of transfer. For example, whereas <ll>-[j] resulted in the lowest rate of transfer (0.01%), <v>-[b] and <d>-[ð] resulted in the highest rates of transfer (99% and 92%, respectively) in the orthography at training condition. The results suggested that the relative degree of acoustic-phonetic salience between an L2 and an L1 sound determines the rate of L1-based transfer (Rafat 2011, 2016). Rafat (2011) also reported that combination productions for <ll>-[j] had been attested in the data, and attributed this to a process akin to the McGurk effect, although a quantitative analysis of this type of error was not conducted.

A different type of acoustic-orthographic integration related to the effect of orthography has been found during the production of Spanish assibilated rhotics ([r] with a sibilant quality/hissing sound), when naïve English-speaking learners are exposed to both auditory and orthographic input at training (Rafat 2015). Participants were assigned to two groups based on input: auditory only and auditory-orthographic. At training, participants in both groups heard auditory stimuli produced by a Mexican speaker of Spanish, whose rhotics were assibilated (e.g., <ahitar> [aitaɾ]). While the auditory-only group participants were exposed only to auditory words accompanied by their images at training, participants in the auditory-orthographic group were exposed to both auditory and orthographic stimuli. Auditory stimuli and their corresponding images were accompanied by written words, which included the grapheme <r> (e.g., <ahitar>) in the auditory-orthographic group. At testing, participants in both groups were shown images of the words and asked to name them. Whereas assibilated rhotics were for the most part produced as sibilants such as [s] and [ʃ] (e.g., [aitas] and [aitaʃ]) when learners
were only exposed to auditory L2 speech, exposure to the grapheme <r> in the auditory-orthographic group promoted both the production of assibilated rhotics [aitaɾ] and approximant rhotics [aitaɹ]. The acoustic and orthographic cues thus interact in different ways, resulting either in the production of assibilated rhotics or English approximant rhotics. First, rhoticity is the less salient feature compared to assibilation in assibilated rhotics, and exposure to the grapheme <r> enhances the less salient feature in the input, leading to target-like productions. Based on an acoustic analysis of the degree of assibilation of the individual tokens in the input, together with the results in the auditory-orthographic condition, the author proposed that the degree of robustness of assibilation in the input modulates orthographic effects. That is, the more salient the degree of assibilation in the input, the more likely exposure to the grapheme <r> at training will lead to the production of an assibilated rhotic by the learners at testing. In the absence of strong assibilation in the auditory input, exposure to orthography at training may override the input and result in transfer, or it might create a “perceptual illusion” of rhotic features, leading to approximant rhotic productions.

That orthography can lead to perceptual illusion has previously been proposed with respect to L1 processing (Hallé, Chéreau, and Segui 2000). Using a phoneme-monitoring task in French, the authors examined the effect of orthographic and phonological incongruence on the perception of [b] and [p] in French-speaking adults. Because of voicing assimilation in French in words such as <absurd> (/bs/ and /bt/ words), the underlying /b/ written as corresponds to [p] rather than [b] in the prefix {ab-} (e.g., /absyʀd/ written as <absurde> is realized as [apsyɾd]). The authors found that the presentation of words whose orthographic representation and phonetic realizations were incongruent yielded a higher detection rate of [b] than [p] in <absurd>-[apsyɾd]. Hallé, Chéreau and Segui (2000) attributed the results to a “perceptual illusion” effect which overrides the input.

Auditory-orthographic interaction may also result in the production of a sound that is not identical to either the L1 or the L2 sound but rather exhibits characteristics of the L1 sound and approximates the L2 sound. A study on Polish-speaking learners’ perception and production of German vowels found that learners produced the German /e:/, which is written in German as <e>, as a different sound, namely a diphthong [ɛe] (Nimz 2016). The grapheme <e> corresponds to /e/ in Polish, but it is acoustically closer to /i/. The author explained the diphthongization by proposing that the learners incorporate both the orthographic and perceptual interferences by starting with an orthography-induced /e/ and satisfy the auditory input by moving towards the quality of the higher vowel /i/ (e.g., [ɛe]).

Rafat (2011, 2015, 2016) and Rafat and Stevenson (2018) provided evidence for the effect of orthography on L2 speech learning of naïve English-speaking learners of Spanish and at the same time proposed that L2 speech learning is a multimodal event despite the fact that previous L2 speech learning models have been based on auditory input only. Although the body of literature has expanded on the effect of orthography on L2 speech learning, there is not much on the effect of orthography on either the L2 speech learning of Persian or heritage speech learning of the Persian language. Persian is an alphabetic language whose script is written from right to left. The modern Persian alphabet is based on the Arabic alphabet with four additional letters. Persian orthography can be considered deep/irregular because some vowels are not marked in writing. These vowels are only marked with diacritics in order to help children learn to read in primary school. Another difficult aspect of the orthography is that several graphemes can map on to the same phoneme. Moreover, Persian orthography is difficult to learn at the initial stages of acquisition because of its particular features such as the number of dots cuing differences in sounds. Therefore, the Persian orthographic system lends itself well to studies on orthographic effects in L2 and heritage speech. L2 learners heavily rely on text
in the classroom setting. As for heritage speakers, a subset of them may not have knowledge of the Persian orthographic system because of lack of instruction, and as a result this may lead to their speech processing, perception and production exhibiting parallels with the migrant L2 learners of English with low or no literacy in their L1 (see Haznedar, Peyton, and Young-Scholten 2018). On the other hand, it is predicted that heritage speakers who receive instruction in Persian will not be immune to auditory-orthographic effects in their speech processing, perception and production.

It still remains a question as to how exposure to two or more orthographic systems, such as the Persian and English orthographies in the case of North Americans, may affect the L2 acquisition and development of heritage phonology. It is predicted that exposure to two or more different orthographic systems may affect the heritage speakers’ processing, perception and production in complex ways.

In addition to the effect of orthography, heritage language learning should be considered a multimodal event and the effect of facial cues should also be taken into account when examining the acquisition Persian as an L2 and Persian as a heritage language. In the case of English and Persian, these two languages differ in terms of the degree of lip rounding and jaw aperture. Moreover, there are different cultural norms around gazing at the interlocutor. Therefore, in addition to perceptual and articulatory constraints previously discussed in models of L2 speech learning, it would be crucial to determine the degree of influence of and reliance on auditory, facial and orthographic cues on how Persian is learned as an L2 or a heritage language or may evolve over time across different generations. Although we have discussed L2 and heritage speech learning from a cognitive perspective, future models can also consider integrating social factors (Swiderski and Rafat 2019).

### 3.4 Heritage speech, language change across generations, and parallels with L2 speech learning

Although heritage speech has been mostly considered from a synchronic point of view and been compared to L2 speech, some studies have also looked at it from a more diachronic perspective, where they have examined sound change across generations in contact situations. This view has been adopted in light of the fact that languages change as a result of contact and thus phonetic and phonological categories merge in bilingual speakers. There is abundant evidence on the attrition of different aspects of the L1 as a result of contact with another language (for a more detailed discussion see Köpke and Schmid 2004). Specifically, the research focusing on L1 phonetic attrition in bilinguals has been growing (e.g., Celata and Cancila 2010; Flege 1987; Guion 2003; Major 1992; Mayr, Price, and Mennen 2012). Phonetic drift in L1 towards the L2 sounds is evidenced in temporal (e.g., Chang 2012; Flege 1997; Major 1992) and spectral aspects of consonant production (Chang 2012; Peng 1993; Ulbrich and Ordin 2014), vowel production (Chang 2012; Baker and Troimovich 2005; Flege 1987; Guion 2003), consonant perception (Celata and Cancila 2010), and intonational features (Mennen 2004).

Flege (1997) was one of the first studies to provide evidence of assimilation of the L1 and L2 phonetic categories. He found changes in the Voice Onset Time (VOT) of French-English and English-French adult bilinguals, where VOT values for French /t/ for both groups were longer than those of their monolingual counterparts. On the other hand, VOT values for English /t/ were shorter than their average native values, again for both groups. Likewise, the second formant frequency (F2) for the vowel /u/ was lower than their native French counterparts for the French group but not for the English group. However, /y/ was produced in a native-like manner by the participants. The results confirmed the predictions that /u/ and /t/ would
be classified as sounds in phonetic categories that already exist in the L1 and /y/ as a sound that is different from an existing category in the L1. Major (1992) also examined VOT values in Brazilian-English bilinguals in the U.S. and similarly to Flege (1997) found evidence of mutual L1-L2 interaction, supporting Flege’s (1995) Speech Learning Model.

VOT drifts in /p,t,k/ in L1 of bilinguals have also been examined from a sociolinguistic point of view in Hrycyna, Lapinskaya, Kochetov and Nagy (2011). A drift towards English VOT values was reported in successive generations (first, second and third generation) of Italian-, Russian-, and Ukrainian-English bilingual communities. They also reported differences between the language groups and suggested that social factors, such as (i) the cohesiveness of a community, which would suggest having enough opportunity for casual speech, (ii) the size of a community, and (iii) attitude towards a particular variety of a language may be responsible for the between-group differences. For detailed information about the characteristics of Persian-English interlanguage and code-switching, read Chapters 26 and 27 in this volume.

Although VOT remains one of the best-studied phenomena in studies that have examined the bidirectionality of language influence on speech production, recently there has been a growing interest in examining a phonetic shift in other aspects of L1 in bilingual speakers. De Leeuw, Mennen and Scobbie (2012) examined the change in the production of the lateral phoneme /l/ in the L1 German of late German-English bilingual speakers living in Canada. They found that the F1 and F2 values of the German /l/ of their bilinguals differed from their native German counterparts and showed a shift towards English. Furthermore, there was a high degree of variability both within and between bilinguals, and not all the participants exhibited this change. They proposed a dynamic system theory: maturational constraints cannot be the only cause of attrition, and various predictors which influence language development in individuals must be considered.

Celata and Cancila (2010) investigated the perception of the geminate-singleton contrast in native speakers of Lucchese Italian and among first generation late Lucchese Italian-English bilinguals (those who emigrated to the U.S.) and second generation/heritage speakers of Lucchese Italian bilinguals (those who were born in the U.S.). The results of a real word and a nonce word identification task revealed that bilingual speakers are significantly worse than the control Lucchese monolingual speakers at the perception of the geminate-singleton contrasts. In particular, the second-generation group exhibited a higher degree of attrition than the first-generation group. Therefore, the authors concluded that the perception of the length contrast has become progressively impaired in their bilingual groups. Given the scarcity of evidence of attrition in bilingual speech at the phonological level, and the fact that gemination had not been previously examined in production studies of phonetic or phonological attrition in these languages, Rafat, Mohaghegh and Stevenson (2017) examined the attrition of L1 geminate-singleton length contrast in Persian-English speaking bilinguals living in Canada. The main goal of their study was to determine whether the geminate-singleton consonant length contrast attrites across three different generations of Persian-English-speaking bilinguals living in Canada. The secondary aim of the study was to draw parallels with L2 speech and shed light on the role of universal phonetic factors on the process of geminate-singleton length contrast attrition in the same population. Previously, Sorianello (2015) had examined the effect of manner/class and voicing on the production of Italian geminates by L2 learners. Rafat, Mohaghegh and Stevenson (2017) examined the effect of manner/class of sounds and voicing as predictors of geminate attrition in eight Persian-English-speaking bilinguals living in Toronto forming three categories of generations: first generation, 1.5 generation and second generation. The 1.5 generation category distinguishes children of Iranian immigrants who had acquired Persian as their first language and came to Canada between the ages of five to fourteen from second
generation heritage speakers of Persian. The productions of the bilinguals were compared with the productions of three homeland variety controls. A word-naming task was conducted. Using Praat software, data were acoustically analyzed. Attrition was defined in terms of changes in mean duration of geminates relative to their singleton counterparts, percentage of geminate-singleton degemination, and category overlap. Results showed that geminates attrite across different successive generations. Moreover, there was some evidence to suggest that geminate realization across generations patterns with typological patterns previously reported, showing that universal phonetic principles such as aerodynamic constraints/articulatory difficulty and acoustic/perceptual salience also constrain geminate realization in bilingual Persian-English speakers. However, there was no evidence to suggest that more marked geminates suffer a higher degree of attrition. This was the first study to examine the attrition of a typologically marked contrast, which considers the role of universal phonetic principles, and markedness in an understudied bilingual community across different generations. This study was later replicated by Alkhudidi, Stevenson, and Rafat (2020), where they examined the phonological attrition of the Arabic geminate-singleton contrast (e.g., /ħamaːm/ “pigeon” vs. /ħamːaːm/ “bathroom”) in the speech of native speakers of Arabic who acquired English after puberty with late bilinguals and heritage speakers. Similarly, another goal of the study was to investigate whether universal phonetic/acoustic factors had an effect on the degree of attrition across generations. Participants performed a delayed word repetition task, which tested the production of geminate and singleton words. Results show that late bilinguals and heritage speakers exhibit reduction in duration in their production of geminate consonants when compared with the monolingual group. Similar to the findings of Rafat, Mohaghegh, and Stevenson (2017) no effect of manner of articulation was found, yet there was an effect of voicing, where more voiced geminates showed a higher degree of attrition across both groups. It is plausible that with a larger sample size, an effect of both manner and voicing could be found with respect to geminate attrition, and the authors recommended further investigation of the effect of universal phonetic factors on phonetic and phonological change in immigrant communities, including heritage speech and across generations.

Apart from the potential effects of voicing and manner of articulation, if the same factors that constrain L2 speech learning also constrain heritage speech, then position in the word should also impact heritage speech production. That equivalence classification is position-sensitive as previously mentioned was first proposed in Flege’s SLM. This has also been shown for Mandarin-speaking learners of Persian (e.g., Falahati 2015) and Persian-speaking learners of Spanish (Rafat 2008). Falahati (2015) investigated the non-native production of rhotics by Mandarin speakers learning Persian as a third language. A series of informal interviews were conducted to collect the data. This resulted in 1252 tokens used for the analysis. The results of an acoustic analysis showed that all speakers produced the allophonic variant trill, which exists in Persian but is absent in both Mandarin and English as their L1 and L2. However, their contextual distribution differed from the native speakers. For more research on the acquisition of segmental and suprasegmental features in Persian as a second/third language, read Chapter 2 in this volume.

Although there are currently no studies that have examined the effect of position in heritage speech in Persian, Cornwell and Rafat (2016) investigated the effect of position in the word on the production of /θ/ and /ð/ by three groups of English speakers in the community of Norwich, Ontario, Canada: English monolinguals, heritage Dutch speakers (early bilinguals), and L1 Dutch/L2 English speakers (late-learning bilinguals). /θ/ and /ð/ productions were measured in both naturalistic and reading tasks. Heritage Dutch speakers produced [θ] and [ð] at similar rates to Monolingual English speakers, but the two groups exhibited different allophonic
Heritage versus second language phonology

realizations, especially when /ð/ was word-initial and /θ/ was word-medial. The findings suggested that despite their ability to produce [θ] and [ð], Dutch heritage speakers may manipulate the inherently variable English /θ/ and /ð/ production to communicate their Dutch cultural identity. This was the first study to examine both heritage Dutch bilinguals in Canada.

Although the studies reviewed previously are important because they highlight the fact that languages change as a result of contact and show parallels between heritage and L2 speech, they also show that heritage speech can be studied in the context of sound change across generations. Therefore, the study of heritage speech production or perception can be conducted from a variationist point of view in order to better gain an understanding of sound change in immigrant communities. That is, in addition to the fact that similarly to L2 speech, universal phonetic factors may constrain heritage speech production, it is important to consider that both individual extra-linguistic factors may also play a role in heritage speech production and language change.

3.5 Child heritage phonology

Very little work has been done on child heritage phonology in comparison with adult L2 speech learning or adult heritage phonology. This is partly due to the difficult nature of eliciting speech from children. Some of the questions that received attention in child bilingual phonology have been acceleration, deceleration, transfer, and whether bilingual phonology develops as one or two separate systems.

With respect to Persian-English bilingual child phonological development, Keshavarz and Ingram (2002) have also addressed the issue of whether bilingual children begin phonological acquisition with one phonological system or two. Analyses of data from a longitudinal study of a Persian-English bilingual infant, Arsham, supported the hypothesis that the child had acquired two separate phonologies with mutual influence; that is, he made occasional use of phonological features of Persian in English words and vice versa. They suggested that this was due to the pattern of exposure to the two languages, and that other children may show a different pattern, depending on their exposure to the two languages and the role of language dominance.

Fakoornia (2017) is another study that is novel in its approach because it considers the development of both the Persian and English phonology of a newly arrived child heritage speaker from Iran to Canada. This study addresses L1 phonetic attrition, and L2 phonetic acquisition of a 9-year-old Persian-English bilingual child. The study aimed to investigate how the manner of articulation of the rhotics, and the two correlates of stress, F0 peak and syllable duration, may change in both the L1 and L2 speech of a Persian-speaking newcomer to Anglophone Canada within a period of one month. A picture-naming task in both Persian and English was carried out in two sessions. Whereas rhotics are approximants in English, in Persian they have a number of different allophonic realizations in different positions. English and Persian also differ in terms of stress realization (Rafat, 2010). In English, stress in bi-syllabic nouns is often on the first syllable, whereas in Persian stress in nouns is on the final syllable. The results of the acoustic analysis revealed that in the second session the number of approximants in Persian in most positions increased, providing evidence for L1 attrition in Persian, and the stress in Persian words was misplaced on the first syllable, providing evidence for influence of English. Yet, as opposed to syllable duration, the F0 peak was not a consistent factor in determining the change in stress pattern. With respect to English, the majority of rhotics was realized as approximants in both sessions. Moreover, accuracy on the duration of syllables and the location of F0 peak in English increased, resulting in producing more English-like
tokens. Thus, it was concluded that the child was acquiring English L2 phonology at the same
time that her Persian was attriting. The findings of this study are novel and contribute to our
understanding of attrition and L2 acquisition in child phonology.

These two studies are unique because they consider heritage child phonology as opposed to
adult heritage phonology, which is most commonly reported. Particularly, they are longitudi-
nal studies. Longitudinal studies are generally rare because they are more difficult to conduct.
For further research on child language acquisition, read Chapter 4 in this volume.

3.6 Conclusion and recommendations

This chapter has contextualized some of the studies on the L2 acquisition and development
of Persian as a heritage language by considering previous work on heritage speech, second
language speech learning patterns, and sound change across generations. In particular, it has
attempted to provide both a multimodal and a sociophonetic perspective on heritage speech
learning, drawing parallels with L2 speech learning. Moreover, it has pointed out the absence
of a model for heritage speech and highlighted some of the issues that have been investi-
gated with respect to heritage and L2 speech. Given the scarcity of literature on this topic,
future work can focus on a variety of issues. Namely, the segmental and prosodic aspects of
Persian as a heritage language in addition to contact with English is important. Other work
can examine other contact situations such as contact with tone languages and languages that
have a non-alphabetic orthographic system such as Mandarin. Moreover, given that speech
learning including L2 speech learning is a multimodal event, it is advisable to consider all
different modalities of speech when investigating the acquisition and development of Persian
as a heritage speech. Furthermore, future studies can focus on both production and percep-
tion. The field would also immensely benefit from longitudinal studies that would shed light
on how heritage language evolves over time and investigate the role of individual differences,
social factors and attitudes, and the type of input heritage speakers are exposed to. Finally,
we would gain a better understanding of how heritage speakers’ phonetics and phonology
develop and may change if we examine both the heritage speakers’ Persian and the majority
language at the same time.

References

Bassetti, B. 2007. “Effects of Hanyu Pinyin on Pronunciation in Learners of Chinese as a Foreign Lan-
guage.” In *The Cognition, Learning and Teaching of Chinese Characters*, edited by Andreas Guder,
Bassetti, B. 2017. “Orthography Affects Second Language Speech: Double Letters and Geminate Pro-
1835–1842.
Benmamoun, E., S. Montrul, and M. Polinsky. 2013. “Heritage Languages and Their Speakers: Opportu-
Heritage versus second language phonology


