Part I

Questions and frameworks for the study of second language speech

1

An introduction to second language speech research

If you are reading this book, you are likely a second/third language learner curious about her or his pronunciation, a linguist, psychologist, or other researcher interested in issues of bi/multilingualism and learning with a particular focus on speech, or both. Regardless of your interests, there are likely to be various aspects of your own speech or that of non-native speakers that have struck you and that you wish to understand better. In this book, we will investigate what is commonly referred to as **foreign accent**, the various differences between the speech perception and production of non-native and (sometimes idealized) native speakers. Our goal is to provide you with an understanding of the principles and phenomena of second language (L2) speech perception and production, theories that have been developed to model these, and the experimental methodologies used to investigate both **segmental** (i.e., consonants and vowels) and **prosodic** phenomena (e.g., lexical stress or tone, rhythm, intonation, and fluency). Our hope is that, once you have finished reading this book, you will either have answers to your questions from previous research or, perhaps more importantly, the ability to conduct your own studies.

In this first chapter, we lay the foundation for the discussion in the entire book, focusing on the major theoretical and empirical questions that guide L2 research, particularly as concerns the acquisition of phonetics and phonology (§1.3). Each theme will be introduced via a series of research questions to be explored throughout the book followed by the presentation of an illustrative study. However, before discussing these central themes of L2 speech research, we first examine the basic structure of speech (§1.1) and a number of concepts relevant to any study of L2 acquisition (§1.2).

1.1 The structure of speech

Human speech is the focus of two major branches of linguistics, namely **phonetics** and **phonology**. Traditionally, these two subdisciplines have been distinguished in

terms of their orientation to the study of spoken language. Phonetics investigates the physical aspects of sound. These include its production (articulatory phonetics), transmission (acoustic phonetics), and perception (perceptual phonetics). Phonology, in contrast, focuses more on the abstract organization of sound systems such as the ways in which sounds may be contrastive and convey meaning (e.g., minimal pairs such as English *pay* /pe/ and *bay* /be/; French *frais* /fʁɛ/ 'fresh' and *vrai* /vʁɛ/ 'true'; and Spanish *gata* /gata/ 'cat' and *rata* /rata/ 'rat'). Phonology also interacts with other linguistic domains including morphology (word structure), syntax (sentence/utterance structure), semantics (meaning), and pragmatics (socially and context-appropriate language use) observed with phenomena including stress and intonation. In the framework adopted in this book, namely that of **experimental phonology**, phonetics and phonology are seen to a great extent as intricately interwoven.

When analyzing both speech perception and production, phoneticians and phonologists normally distinguish between two levels of organization, namely segments and prosody. The study of segments focuses not only on the individual realization of consonants and vowels but also on their production in sequences, that is on their **coarticulation**. In contrast, prosody refers to all phenomena that serve to group individual sounds into larger units including syllables, metrical feet relevant for the assignment of stress, and even larger units including phonological phrases and utterances necessary for the organization of phenomena such as intonation; prosody also looks at fluency phenomena including rhythm and timing. In Part III of this book, each chapter will begin with a thorough discussion of the phonetics and phonology of vowels (Chapter 4), consonants (Chapters 5, 6), sequences (Chapter 7), or prosody (Chapter 8) necessary for understanding and undertaking L2 speech research. At the end of this chapter, you will also find some useful readings if you wish to explore further these two subdisciplines of linguistics.

1.2 Core concepts and terminology in second language acquisition

Within the larger field of bilingualism and multilingualism, **second language (L2) acquisition** refers to the acquisition of a language following the (mainly complete) acquisition of one's first language (L1). In those cases involving child learners where acquisition begins before the age of seven to ten years and results in very proficient mastery of the target language, researchers often refer to **child** or **early L2 acquisition**. Unless otherwise highlighted, the research discussed in this book focuses on the former, more common type of L2 acquisition.

In more recent years, researchers have begun to distinguish clearly between L2 acquisition and **third language (L3) acquisition**. This latter type of learning involves the acquisition of a non-native language subsequent to that of another

non-L1 (see De Angelis (2007) for a general overview and Wrembel, Gut & Mehlhom (2010) for a collection of recent studies on L3 phonology in particular). With the exception of this more recent research, researchers have most often not distinguished between L2 and L3 acquisition. Accordingly, in this book, with the exception of those instances where we discuss studies that seek to investigate L3 acquisition in particular including the effect of a previously acquired L2 on learners' perception and production, we will adopt the more common approach of referring to all such learning as L2 acquisition.

In L2 acquisition, learners begin with the knowledge of their L1 as well as of any other previously (partially) acquired non-native languages, and then move gradually towards the language being acquired, or **target language (TL)**. Acquisition is driven by the ambient spoken and written language, that is, the **input** to which learners are exposed. Over the course of acquisition, learners' perceptual and productive knowledge of the target language will be represented cognitively as a series of **interlanguage (IL) grammars**. The term *interlanguage* (e.g., Corder, 1967; Selinker, 1972; Adjemian, 1976) refers to the fact that, once acquisition has begun, learners' L2 perceptual and production knowledge normally falls somewhere in between that of their L1, due to **cross-linguistic influence (CLI)** of this dominant language, and the target language. Often, in spite of the variability observed between learners of a given target language who share the same L1 as well as with an individual learner at a given point in time, it is possible for researchers to distinguish **developmental sequences**. These are universally observed patterns of acquisition that involve a series of stages characterized by patterns and processes common to all learners regardless of their L1. The transition from one stage of development to the next is gradual, with the rate of change varying from learner to learner based on a number of variables including the quantity and quality of input and a learner's overall aptitude for L2 acquisition as determined by factors including phonological memory and motivation. Furthermore, features of more than one of the stages identified by researchers may be observed in learners' perception or production at a given point in time. If a learning plateau is observed during which acquisition appears to have stopped, even temporarily, one speaks of **fossilization**. At the point at which it appears that acquisition is complete or will advance no further, the IL grammar may be referred to as an **end-state** grammar and the characteristics of this grammar are referred to as a learner's **ultimate attainment**, which may be more or, more commonly, less native-like. In the domain of phonetics and phonology, L2 end-state grammars typically, but not necessarily, differ from those of native speakers. As we discussed at the beginning of this chapter, this is referred to as foreign accent.

In the next section, we will investigate the research questions that specifically target each of these concepts and processes related to the L2 acquisition of speech.

Each of these will be exemplified with some previous research including a series of illustrative studies.

1.3 L2 speech research questions

All research is driven by the desire to answer one or more questions concerning a given phenomenon; work on L2 speech is no exception. In the following sections, we discuss many of the recurrent theoretical themes found in experimental work on L2 speech perception and production. Answers to these questions, at times partial, will be offered in the chapters that follow.

1.3.1 The role of input

Theories of language acquisition assign a primary role to the linguistic input based on which learners construct both perceptual and production grammars. As mentioned earlier, in the case of speech learning by literate learners, input consists of both aural and written language.

The first question that one can identify in previous research is *How does input trigger speech learning?* This question may be of a particular interest in the context of L2 acquisition given that, even when faced with large quantities of native input, learners may fail to acquire native-like perceptual and/or productive competence. How one answers this question depends much on the particular theoretical framework adopted. Two broad theoretical frameworks can be identified. The first is that of researchers who assume that language learning, including L2 speech acquisition, is primarily associative, with learners establishing connections between form and meaning. Such an approach is variably referred to as **functionalist** or **associative-connectionist** (see, e.g., Ellis (2007) for a general discussion of this approach to L2 acquisition). On this approach, input drives acquisition with learners forming generalizations based on the language to which they are exposed. Such generalizations principally involve associations between a form in a given context and its meaning. For example, in the case of the acquisition of the phonemic inventory of a given target language, once learners have begun to be able to segment the speech signal into words, they begin to identify minimal pairs from which they can deduce which phones are contrastive and thus phonemic. The second broad theoretical framework includes researchers who adopt **nativist** models. Such models propose that humans are endowed with a language faculty, which includes innate knowledge concerning the universal set of possible linguistic categories, structures, and principles of their organization. In the domain of phonology, this would include the possible features that may be used to establish phonemic contrasts, minimal and maximal syllable templates, and the basic units and principles used to construct

the representations of stress systems. The most commonly encountered work within this framework is conducted within Chomsky's Universal Grammar (see White (2003) for detailed discussion of this framework for the L2 acquisition of morphosyntax and Archibald (1998) for its use in modelling the acquisition of L2 phonology). Within nativist models, input serves to trigger language-specific settings of parameters governing the wellformedness of linguistic representations or, in constraint-based frameworks, to allow for the ranking of constraints that evaluate wellformedness and complexity. The framework within which one undertakes L2 speech research is a function of a researcher's larger understanding of the fundamental nature of human language and, one must admit, the community of researchers within which one was trained and currently works. The overwhelming majority of previous L2 speech research has adopted a functionalist approach.

A second principal question related to the role of input is *What are the consequences of differences in input quantity, quality, and modality for L2 speech learning*? Given the primary importance of input, these three variables each have an enormous potential effect on the type of cross-linguistic influence observed, patterns of development, and learners' ultimate attainment including the degree of intra- and interspeaker variability. The input to which a given group of L2 learners has access may differ greatly in quantity and quality as well as **modality** (i.e., auditory, visual, or written) vis-à-vis the input available to children or other L2 learners. In terms of quantity, children are exposed to vast amounts of auditory native speaker input from birth. This large quantity of ambient language allows child learners to perceive and, with sufficient practice, produce the full range of segmental and prosodic phenomena of the target language being acquired. In contrast, the input available to many L2 learners is often impoverished both in terms of quantity and quality (see Moyer (2009) for an overview). For example, when a target language is encountered principally in a formal learning environment such as the classroom and is not the majority language of the surrounding speech community, learners may be limited to a few hours of input a week from a single individual who may or may not be a native speaker. In a classroom setting, the speech of one's fellow learners also constitutes input to learning, input which is clearly deficient, at least in terms of quality. The consequence of such a situation is that learners may come to acquire perception and production grammars that accurately reflect the input to which they were exposed, although such input may not reflect the speech of native speakers. As mentioned above, differences in input also involve modality. Whereas child learners do not have access to written language until the onset of the acquisition of literacy, most adult learners are literate, at least those who have served as participants in the vast majority of the studies reported in published research (see Young-Scholten, 1995, p. 114 for more discussion). We will return to

the question of the effect of literacy, particularly that of written input, on L2 speech learning in §1.3.6. Within research on the effects of differences in quality and quantity of input, a particular sub-domain studies the potentially facilitative effect of training in the laboratory (e.g., Iverson, Hazan & Bannister, 2005; Motohashi-Saigo & Hardison, 2009) and instruction in the classroom (e.g. Macdonald, Yule & Powers, 1994; Saito & Lyster, 2012). Such work investigates the degree to which various techniques and paradigms including high variability training, acoustic signal manipulation, and form-focused instruction allow L2 learners to more effectively analyze the input when acquiring both segmental and prosodic aspects of the TL.

Illustrative study 1.1. Freed, Segalowitz & Dewey (2004)

Overview: This study investigated differences in gains in oral performance and fluency between students in three learning contexts differing in input quantity and quality: traditional classroom study in the home country (AH – at home), intensive traditional classroom study accompanied by a variety of out-of-class activities in the target language (IM – immersion), and a study abroad (SA) experience combining formal instruction and possibilities for regular interaction with the native language speech community.

Languages: French (TL); English (L1)

Dependent variables: *Oral performance*: (1) Total words spoken; (2) Duration of speaking time; (3) Length (in words) of longest turn; *Oral fluency*: (1) Speech rate; (2) Hesitation-free speech runs; (3) Filler-free speech runs; (4) Fluent runs versus repetition-free speech runs; (6) Grammatical-repair-free speech runs.

Independent variable: *Learning context*: The three contexts differed in terms of the total number of hours of formal instruction on average (AH: 12 weeks, 36–48 hours; IM: 7 weeks, 123 hours; SA: 12 weeks, 197 hours) and opportunities for contact and use of French outside of class (AH: highly limited; IM: a variety of cultural and other activities throughout the period with students taking a 'language pledge' to communicate only in French including outside of class; SA: the programme took place in Paris with the majority of students participating in courses offered by their home institution)

Research questions: (1) Are there salient differences in the acquisition of oral fluency by students who have studied abroad, when compared to students whose learning takes place in IM programmes or the regular AH language classroom? (2) Do time-on-task factors (e.g., instructional time, out-of-class time spent interacting orally with native speakers or using the language within the literate domain) vary in each of these contexts? (3) To what extent are the measured differences in oral fluency associated with these time-on-task features? (p. 280)

Methodology

Participants: 28 adults (mean age = 21.3 years) studying in an American university
Tasks: (1) Oral interviews: 15–30 minutes in length, conducted both preceding and following the period of instruction; (2) Questionnaire: a language contact profile designed to determine language use (how much time was spent reading, listening, writing, and speaking and in which language) and interaction (with whom).
Data analysis: *Oral interviews*: For each participant, two one-minute samples from the pre- and post-course interviews were transcribed and verified by two sets of two individuals and then analyzed for the nine dependent variable temporal and hesitation phenomena. Fluency gain scores were calculated by subtracting pre-course from post-course scores.

Main findings

(1) Whereas the students in the three groups differed minimally on their pre-course measures, statistically significant gains were made only by the IM group and only for a subset of the measures of oral performance (total number of words, length of longest turn) and fluency (rate of speech). Moreover, for a composite fluidity variable (combined hesitation-free, filler-free, repetition-free, grammatical-repair-free, and fluent speech runs), the IM group made the most overall gains in fluency follow by the SA group; the AH group made no gains at all.
(2) The IM group reported the greatest number of hours of speaking, reading, writing, and listening to French.
(3) The only unambiguous, significant correlation discovered existed between the composite fluidity variable and time spent on writing.

1.3.2 Cross-linguistic influence

A widely observed phenomenon in any bi- or multilingual speaker is the interaction of the individual's various languages in both comprehension/perception and production. This phenomenon, alternatively referred to as cross-linguistic influence (CLI), **transfer**, or **interference**, occurs in L1 bilingual language acquisition (e.g., Paradis, 2001; Fabiano-Smith & Goldstein, 2010), L2 acquisition (e.g., Flege & MacKay, 2004; Lee, Guion & Harada, 2006), and adult balanced bilinguals (e.g., Fowler, Sramko, Ostry, Rowland & Hallé, 2008). In the case of L2 phonological acquisition in particular, CLI has been observed to influence individual segments (e.g., McAllister, Flege & Piske, 2002; Aoyama, Flege, Guion, Akahane-Yamada & Yamada, 2004) as well as sequences including coarticulation (e.g., Levy & Strange, 2008; Zsiga, 2003), and all aspects of prosody (e.g., Steele, 2002 for syllable structure; Matthews & Brown, 2004; Archibald, 1997; Dupoux, Pallier,

Sebastián-Gallés & Mehler, 1997 for stress; Hallé, Chang & Best, 2004; Francis, Ciocca, Ma & Fenn, 2008 for tone; Aoyama & Guion, 2007; Nava & Zubizaretta, 2009 for rhythm; and Mennen, 2004; Swerts & Zerbian, 2010 for intonation).

One of the most obvious cases of CLI is a speaker's use of L1 categories or prosodic patterns in their L2 production. For example, French-speaking learners of English may realize the voiceless interdental fricative /θ/ of English words such as *think*, *athlete*, and *bath* as well as the voiced interdental fricative /ð/ of *this*, *although*, and *bathe* with their L1 /s/ and /z/ respectively (i.e., as [s]ink, [z]is) given the absence of the former consonants in French. Such substitutions are specific to a learner's particular L1 **variety** (i.e., dialect). Whereas the substitution pattern above holds true for European French speakers, native speakers of Quebec French substitute rather /t/ and /d/ (i.e., [t]ink, [d]is; e.g., Brannen, 2002; Picard, 2002; Trofimovich, Gatbonton & Segalowitz, 2007). Native dialect effects are also observed in perception. A variety of recent studies (Chládková & Podlipský, 2011; Escudero, Simon & Mitterer, 2012; Escudero & Williams, 2012; among others) have demonstrated that L2 listeners categorize TL vowels based not on the phonemic categories of their first language but rather the acoustic categories of their particular L1 variety. We will examine the effect of L1-based CLI on perception further in Chapter 2 when discussing theories of L2 speech perception.

At the prosodic level, the effect of L1-based CLI has been studied most often as concerns its consequences for syllable structure modification and misplacement of lexical stress. In terms of the former, learners may modify TL consonant sequences that exceed the complexity permitted in their L1, most often via vowel epenthesis (e.g., Broselow, Chen & Wang, 1998; Abrahamsson, 1999) or consonant deletion (e.g., Eckman, 1987; Hancin-Bhatt & Bhatt, 1997). For example, languages like English, French, and Spanish allow initial stop-liquid clusters (e.g., *play* /ple/, *drain* /dren/; *clouer* /klue/ 'nail (INF)', *privé* /pʁive/ 'private'; *globo* /globo/ 'balloon', *crema* /krema/ 'cream') not permitted in languages whose dominant syllable template is CV (consonant-vowel) or CVC (consonant-vowel-consonant) including Mandarin and Cantonese. L2 learners with such latter L1s may realize TL stop-liquid clusters via the insertion of a vowel (e.g., English *drain* /dren/ realized as [dəɹen]; Spanish *globo* /globo/ realized as [gəlobo]) or by deletion of one of the consonants, most commonly the liquid (e.g. English *play* /ple/ realized as [pe]; French *privé* /pʁive/ realized as [pive]). Epenthesis and deletion both serve to bring TL syllable structure into conformity with a learner's L1; we will discuss syllable structure modification in further detail in §7.4. In the case of lexical stress, CLI results in the application of L1 stress patterns to TL lexical items (e.g., Archibald, 1993a; 1997). In many cases, this results in the misplacement of stress. For example, a Spanish-speaking learner of English, whose L1 favours penultimate stress, might pronounce English *afford*, *concentrate*, *currently*, *innocent*, and *retrieve* stressing