
INTRODUCTION

1.1. AIM AND THEORETICAL FRAMEWORK

The aim of this book is to describe the phonological system of Portuguese and to discuss its functioning in the light of recent phonological theories. Chapters 1 and 2 are introductory: they describe some of the phonetic, phonological and morphological characteristics of European Portuguese (henceforth EP), where relevant including data from Brazilian Portuguese (henceforth BP). Chapter 2 also deals with theoretical aspects such as feature geometry and underspecification. Some of this material will be further discussed in later chapters. From Chapter 3 onwards, the analyses are based on autosegmental theory, and Portuguese phonological processes, both segmental and prosodic, are set out in terms of a multilinear structural organization. The discussion of nasalization, phonological and morphological processes, syllable and stress presupposes such a theoretical framework. All the models we use are from the same theoretical background and have been chosen for their explanatory power.

1.2. RELEVANT HISTORICAL MILESTONES AND SYNCHRONIC VARIATION

Portuguese is a Romance language closely linked to Castilian and Catalan. In Europe, Portuguese is spoken in Portugal. In traditional studies on dialectology, and for historical reasons, the dialects spoken in Galicia (Spain) have been grouped with the Galician-Portuguese dialects (see Cintra, 1971a). We shall not consider them here. It was in Galicia, which was founded in the third century and was a part of the Roman-occupied lands of Gallaecia and Asturica, that the Portuguese language emerged from Vulgar Latin. The inhabitants of this peripheral and inaccessible area retained in their language the typical characteristics of the archaic Low Latin spoken by their colonizers.

The languages spoken on the Iberian Peninsula during the Roman occupation influenced the phonological differentiation between the Hispanic languages and dialects. According to authors such as Baldinger (1958), this influence, for example, suppressed the initial etymological [f] in Castilian and the fricative pronunciation of the Latin plosive [b], which was common to both the northern

Portuguese and the Castilian dialects. Between the fifth and the seventh centuries, the Peninsula was invaded by Germanic peoples, and a Suebic kingdom was established in the northwest of the Peninsula. The region became increasingly isolated so that the type of Latin spoken there grew to be unique. When the Visigoths overthrew the Suebic rulers in the seventh century, some of the phonological characteristics that set Galician-Portuguese apart from Castilian had already formed. The intervocalic pronunciation of Latin [n] and [l] had disappeared and the Latin clusters [pl], [kl] and [fl] had become palatalized. At the same time, the classical Latin vowel system had reached the end of its progression with the loss of quantity opposition and, in Portuguese, there was a replacement of the short [e] and [o] vowels by the corresponding open vowels [ɛ] and [ɔ], although no diphthongization occurred, as happened in Castilian.

Earlier texts written in Portuguese date back to 1175—*Notícia de Fiadores*—and 1214–16—*Notícia do Torto* and the *Testamento de D. Afonso II*.¹

The spread of the Portuguese language kept pace with the conquest of Portugal from the Moors during the thirteenth century, extending right down to the province of Algarve in the extreme south of the country. Later, Portuguese spread to every continent in the world and at the end of the sixteenth century and during the seventeenth, apart from being spoken in Brazil, it served as a *lingua franca* along the West-African coast and in the ports of India (Indo-Portuguese) and South-East Asia (Malay-Portuguese).

Today, Portuguese is the national language of Portugal and Brazil. It has been made the official language of Angola, Mozambique, the Cape Verde islands, Guinea-Bissau, and the São Tomé and Príncipe Islands. In Asia, only Macao has officially kept Portuguese, which is also spoken in the eastern part of the island of Timor. Large communities of emigrants keep Portuguese alive in North America and in various countries in Europe.

The two most important varieties of Portuguese, which define so-called educated standards of the language, are European Portuguese and Brazilian Portuguese. The most obvious differences between these two varieties are located in the unstressed vowel system—the vowels are more audible in BP than in EP.

With regard to dialects, there are no striking differences between the dialects in either Portugal or Brazil. In European Portuguese there are two main groups of dialects: northern Portuguese and central-southern Portuguese. The distinction between these groups lies mainly in the fricative consonant system, traditionally called ‘sibilants’: the northern dialects have kept their Latin apico-alveolar fricatives, as is also the case in northern Castilian, and in some areas, these consonants coexist with the dental ones, so reflecting the distinction shown in the written form (*passo* ‘step’, an apico-alveolar, and *paço* ‘palace or court’, a dental); in the central-southern dialects these fricatives have been replaced by the dental ones although the different forms of spelling have been kept. The dialects on the archipelagos of Madeira and the Azores, while they have their own peculiarities, share the general characteristics of the central-southern dialects.

¹ See Cintra (1971b), Costa (1977), and Martins (1999).

According to Nascentes (1953), there are two groups of dialects in Brazilian Portuguese: the ones from the north (the Amazon, the north and the north-east) and those from the south (the central and southern regions). The Northern dialects tend to have more open prestressed vowels. There are several linguistic atlases, many monographs and quite a number of general articles available on the Portuguese Brazilian dialects. However, Cunha and Cintra (1984) believe that the sheer vastness of Brazil has made the job of presenting an organized, detailed study of all the existent dialect variations extremely difficult. Even if the variation in dialects is not very great, the enormous variation in sociolects presents the Brazilian school with a formidable challenge.

In the Portuguese-speaking African countries, the number of Portuguese speakers is still fairly low, although it is increasing slowly as more children attend school. The lack of systematic, exhaustive research into the type of Portuguese spoken in these countries has made it impossible to draw up a profile. Nevertheless, in very general terms, it can be said, that the unstressed vowels are more audible in Africa than in European Portuguese.

In Angola and Mozambique, Portuguese coexists with other national languages, mainly belonging to the Bantu group. In Guinea-Bissau, the São Tomé and Príncipe islands and the Cape Verde islands different Creoles have grown from Portuguese. The Creoles spoken in São Tomé and in Cape Verde are national languages and are the only ones standing on an equal footing with Portuguese in these two countries.

1.3. STUDIES ON PORTUGUESE

The first Portuguese grammars date from the sixteenth century: Fernão de Oliveira (1536) and João de Barros (1540). These compilations naturally provide one of the most valuable sources of knowledge about the history of the Portuguese phonology. Barbosa's grammar (1822), written in emulation of Arnauld and Lancelot's *Grammaire Générale et Raisonnée* (1660), is the most important philosophical grammar of Portuguese. Philosophical orientation notwithstanding, it provides insights on the Portuguese prosody.

At the beginning of the twentieth century, several historical grammars were published. They are vital for understanding the development of the phonology of Portuguese both in Portugal and in Brazil. It is worth singling out from among them Said Ali (1921–3), Nunes (1919) and Williams (1938). Pre-theoretical studies on Portuguese phonology are those of Viana (1883), who shows remarkable phonological intuition, and, among others, Vasconcellos (1901), who throws much light on Portuguese dialects.

Extremely full structural descriptions of the phonology of present-day Portuguese are to be found in Barbosa (1965) for European and Câmara (1953 and 1970) for Brazilian Portuguese. Cunha and Cintra's work (1984) looks at both varieties from a traditional stance.

In the last twenty years, work published on the Portuguese language has taken into account standard generative grammar and related theories. In Portugal, phonological studies have been appearing since the mid-1970s, for example, Mateus (1975) and Andrade (1977), who write on the phonology of European Portuguese in the *Sound Pattern of English* model (henceforth SPE), and Callou and Leite (1990), who study the phonology of Brazilian Portuguese.

The Proceedings of the Portuguese and Brazilian Linguistic Associations (APL and ABRALIN) contain some enlightening articles on the current state of research. Many studies in Portuguese phonetics have been published since 1980, e.g. Delgado-Martins on the phonetic study of prosodic facts (1982 and 1983), Viana (1987) and A. Andrade (1987). A few articles are dedicated to applying metrical and autosegmental theories to Portuguese phonology: Andrade (1983), Andrade and Viana (1988a, b), Andrade and Laks (1987, 1991, 1996), Abaurre (1991), Bisol (1989) and Wetzels (1991).

The present study of the Portuguese phonological system is based on European Portuguese and presents data from Brazilian Portuguese either to argue in favour of certain proposals or to explain different phonological processes. European Portuguese data will be taken from the standard dialects spoken in Lisbon and Coimbra, which are accepted in Portugal as a reference for teaching Portuguese as a second language and are the most commonly heard on radio and television. Some remarks about other dialects will be made on appropriate occasions. For Brazilian Portuguese, in which there is more than one standard, the widely diffused dialect spoken in Rio de Janeiro ('carioca') will be referred to.

The most complete monolingual dictionaries of Portuguese (Figueiredo, 1899; Silva, 1948–59; Aurélio, 1986) do not include phonetic transcriptions. Only Vilela (1990), a small dictionary compiled for pedagogical purposes, gives instruction on pronouncing words. Viana (1883) contains some interesting remarks about the pronunciation of vowels and sequences of vowels.

1.4. OVERVIEW OF THE BOOK

The present chapter briefly surveys the historical aspects and synchronic variations of the language which are relevant to an understanding of the contemporary Portuguese. The description of the phonological system included in Chapter 2 illustrates the distribution and the phonetic characteristics of consonants, vowels and glides in Portuguese, employing a wide range of data. We take an autosegmental perspective in the analysis of the segments and propose an organization of the features which accords with the feature geometry model. We also adopt the underspecification view, and we state the underspecification of Portuguese consonants and vowels.

Chapter 3 looks at prosodic structure. After a phonotactic description, the non-linear theoretical model is introduced to account for the behaviour of the

segments with respect to the internal structure of the syllable and the phonological word.

In Chapters 4 and 5 we present the general characteristics of Portuguese morphology—nominal and verbal systems—in the framework of lexical phonology. Chapter 4 concerns inflectional processes in which morphology interacts with phonology to affect the application of several phonological rules. Chapter 5 analyses derivational processes and discusses morphophonological phenomena occurring in derivational processes. Chapter 6 deals with word stress—main, secondary and echo—using the ‘grid only’ model as our analytical tool. The focus of Chapter 7 is on phonological processes that are not related to the morphological structure of the word. Here the rules applied on final syllable consonants and the very peculiar process of nasalization are analysed. We also deal with the reduction of unstressed vowels which occurs in European Portuguese. The last part of the chapter reviews the consequences of applying external sandhi rules.

1.5. PHONETIC TRANSCRIPTION AND CONVENTIONS

The phonetic transcriptions included throughout this work are broad and, up to a point, abstract transcriptions representing a class of individual realizations. Although they vary in phonetic details, these differences are not deemed linguistically relevant. The notation used is that of the most recent version (1993) promulgated by the International Phonetic Association (IPA). Following usual conventions, the phonetic transcriptions are written in square brackets and the underlying representations are enclosed in slashes. Sigma (σ) is the symbol for syllables. Whenever necessary, morpheme boundaries (+) word boundaries (#) and syllable boundaries (– or \$) will be used. Primary stress is indicated by the diacritic (ˈ) usually placed over the stressed vowel.

Italics represent a sound that may be interpreted as being etymological. Orthographic transcriptions are enclosed in angles. Examples that occur in running text are given in *italic* and referred to in their orthographic form. Sometimes a word in its orthographic form will contain one or more sounds represented in phonetic transcription (e.g. am[ó]r).

The lateral [ɭ] used in the EP dialects is a velarized consonant. It is pronounced by raising of the tongue body towards the velum (as in pronouncing the English ‘dark l’).

In the standard varieties of EP and BP described here we seldom find, in word-initial position or between vowels, the typical alveolar trill that implies ‘usually three vibrating movements’ (Ladefoged, 1975). It can be found in some dialects of Portugal and in the south of Brazil. Fairly often, an uvular fricative occurs instead, being represented by the symbol [ʁ]. The alveolar trill, a very short one where there is only a single vibrating movement (almost an alveolar tap), is denoted here by [ɾ]. In some Brazilian dialects it can occur as a retroflex sound. The

symbol [χ] represents a voiceless uvular fricative which is usually pronounced in syllable-final position in BP.

1.6. A WORD ON ORTHOGRAPHY

Portuguese spelling is phonological; it is also fairly conservative. The spelling system of the vowels has single symbols. The five letters, ⟨a, e, i, o, u⟩ represent in fact nine vowels: [a], [ɐ], [ɛ], [e], [i], [ɔ], [o], [u] (and [ĩ] in European Portuguese). The glides [j] and [w] are usually represented by ⟨i⟩ and ⟨u⟩ (but they may also be denoted by ⟨e⟩ and ⟨o⟩). The correspondence between letters and vowels and glides, and the different pronunciation in the two varieties, is given in (1) below:

- (1) ⟨a⟩ → [a] *caso* [kázu] ‘case’
 ⟨a⟩ → [ɐ] *café* [kɛfɛ́] (EP) ‘coffee’
 ⟨a⟩ → [ɐ] *sopa* [sópɐ] ‘soup’
 ⟨e⟩ → [ɛ] *belo* [bélu] ‘beautiful’
 ⟨e⟩ → [e] *seco* [séku] ‘dry’
 ⟨e⟩ → [i] *ermida* [iɾmídɐ] ‘small church’
 ⟨e⟩ → [i] *estar* [iʃtár] (BP) ‘to be’
 ⟨e⟩ → [i] *nove* [nóvi] (BP) ‘nine’
 ⟨e⟩ → [ĩ] *dever* [dívér] (EP) ‘duty’
 ⟨e⟩ → [ĩ] *nove* [nóvi] (EP) ‘nine’
 ⟨e⟩ → [i]/[j] *cear* [siár]/[sjár] ‘to have supper’
 ⟨i⟩ → [i] *vi* [ví] ‘I saw’
 ⟨i⟩ → [i] *tirar* [tirár] ‘to take’
 ⟨i⟩ → [j] *pai* [páj] ‘father’
 ⟨i⟩ → [i]/[j] *pior* [piór]/[pjór] ‘worst’
 ⟨o⟩ → [ɔ] *bola* [bóɫɐ] ‘ball’
 ⟨o⟩ → [o] *força* [fóɾsɐ] ‘strength’
 ⟨o⟩ → [u] *poder* [pudér] (EP) ‘power’
 ⟨o⟩ → [u] *pato* [pátu] ‘duck’
 ⟨o⟩ → [u]/[w] *voar* [vuár]/[vwár] ‘to fly’
 ⟨u⟩ → [u] *tudo* [túdu] ‘everything’
 ⟨u⟩ → [u] *murar* [murár] ‘to enclose’
 ⟨u⟩ → [w] *pau* [páw] ‘stick’
 ⟨u⟩ → [u]/[w] *suor* [suór]/[swór] ‘sweat’

The plosive consonants generally do not have different pronunciations in the two varieties. They are represented by the following letters:

- (2) ⟨p⟩ → [p] *pá* [pá] ‘spade’
 ⟨b⟩ → [b] *boa* [bóɐ] ‘good’ (fem.)

- ⟨t⟩ → [t] *u* [tú] ‘you’
 ⟨d⟩ → [d] *dar* [dár] ‘to give’
 ⟨c^{a,o,u}⟩ → [k] *casa* [kázɐ] ‘house’
 ⟨g^{a,o,u}⟩ → [g] *gato* [gátu] ‘cat’

(i) The letter ⟨q⟩ can be used to represent the sound [k] but it is always followed by ⟨u⟩ which is normally not pronounced (*quero* [kɛru] ‘I want’). The letter ⟨g⟩ placed before ⟨e⟩ and ⟨i⟩ represents a plosive only if it makes a string with ⟨u⟩ which does not need to be pronounced either (*guita* [gítɐ] ‘string’). However, in these two cases the ⟨u⟩ represents a glide in some words, as in *quarto* [kwártu] ‘room’ and in *linguística* [lĩgwíftikɐ] ‘linguistics’, or in BP *questão* [kweʃtɛw̃] ‘question’.

(ii) In some Brazilian dialects, ⟨t⟩ and ⟨d⟩ placed before ⟨i⟩ represent the affricates [tʃ] and [dʒ] as in *tia* [tʃíɐ] ‘aunt’, *dia* [dʒíɐ] ‘day’ (see 2.1.2. below).

The fricative dental and palatal consonants maintain a difference in orthography due to a different pronunciation that has disappeared from most dialects (see 2.1.1 (iv) concerning the different pronunciation of these fricatives, except for voiced palatals, in some dialects).

- | | |
|---------------------------|---|
| (3) ⟨ss⟩ → [s] | <i>passo</i> [pásu] ‘step’ |
| ⟨ç⟩ → [s] | <i>paço</i> [pásu] ‘palace’ |
| ⟨c ^{e,i} ⟩ → [s] | <i>cego</i> [ségu] ‘blind’ |
| ⟨s⟩ → [s] | <i>saber</i> [sɐbéɾ] ‘knowledge’ |
| ⟨s⟩ → [z] | <i>casa</i> [kázɐ] ‘house’ |
| ⟨s⟩ → [ʃ] | <i>pás</i> [páj] ‘spades’ |
| ⟨z⟩ → [z] | <i>gozo</i> [gózu] ‘joy’ |
| ⟨z⟩ → [ʃ] | <i>paz</i> [páj] ‘peace’ |
| ⟨ch⟩ → [ʃ] | <i>chá</i> [já] ‘tea’ |
| ⟨x⟩ → [ʃ] | <i>xá</i> [já] ‘Shah’ |
| ⟨x⟩ → [s] | <i>sintaxe</i> [sĩtási]/[sítási] ‘syntax’ |
| ⟨x⟩ → [z] | <i>êxodo</i> [ézudu] ‘exodus’ |
| ⟨x⟩ → [ks] | <i>táxi</i> [táksi] ‘taxi’ |
| ⟨g ^{e,i} ⟩ → [ʒ] | <i>gingar</i> [ʒĩgár] ‘to roll’ |
| ⟨j⟩ → [ʒ] | <i>jacto</i> [zátu] ‘jet’ |

⟨l⟩ and ⟨lh⟩ represent lateral consonants in syllable-initial position. At the end of the syllable only ⟨l⟩ occurs and it represents the velarized lateral [ɫ] (EP) or the glide [w] (BP) (for more on the distribution of the laterals see 2.2.2(ii)).

- | | |
|---------------|------------------------------|
| (4) ⟨l⟩ → [l] | <i>mala</i> [máleɐ] ‘bag’ |
| ⟨l⟩ → [ɫ] | <i>mal</i> [máɫ] (EP) ‘evil’ |
| ⟨l⟩ → [w] | <i>mal</i> [máw] (BP) ‘evil’ |
| ⟨lh⟩ → [ʎ] | <i>alho</i> [áʎu] ‘garlic’ |

The liquids [r] and [R] are represented respectively by ⟨r⟩ and ⟨rr⟩ except word-initially, where the uvular trill is denoted by one single ⟨r⟩. See (5).

(5) ⟨r⟩	→	[r]	<i>caro</i> [káru] ‘expensive’
⟨r̄⟩	→	[R]	<i>rato</i> [Rátu] ‘mouse’
⟨rr⟩	→	[R]	<i>carro</i> [káRu] ‘car’

⟨m⟩, ⟨n⟩, ⟨nh⟩ represent nasal consonants in syllable-initial position. In syllable-final position (before a consonant or word-finally), ⟨m⟩ and ⟨n⟩ represent the vowel nasalization: ⟨m⟩ before ⟨p⟩, ⟨b⟩ and word-finally, ⟨n⟩ before the other consonants; ⟨nh⟩ does not occur in this position.²

(6) a.	⟨m⟩	→	[m]	<i>mama</i> [mémɐ] ‘mamma’
	⟨n⟩	→	[n]	<i>mana</i> [mẽnɐ] ‘sister’
	⟨nh⟩	→	[ɲ]	<i>manha</i> [mɛɲɐ] ‘slyness’
b.	⟨am⟩	→	[ẽ]	<i>campina</i> [kẽpínɐ] ‘level land’
	⟨am̄⟩	→	[ẽw̄]	<i>falam</i> [fálẽw̄] ‘they speak’
	⟨em⟩	→	[ɛ]	<i>embora</i> [ẽbórɐ] ‘even so’
	⟨em̄⟩	→	[ẽ]	<i>viagem</i> [viázẽ]
	⟨em̄⟩	→	[ẽ]	<i>viagem</i> [viázẽ]
	⟨im̄⟩	→	[ĩ]	<i>assim</i> [ɛsí] ‘so’
	⟨en̄⟩	→	[ɛ]	<i>pensar</i> [pẽsár] ‘to think’
	⟨on̄⟩	→	[õ]	<i>contar</i> [kõtár] ‘to count’

The practice of writing the nasal consonants which merely indicate that the vowel is nasal although they are not pronounced as such is another example of the phonological nature of written Portuguese.

The letter ⟨h⟩ never represents an aspiration: either it represents nothing (*humano* [umẽnu] ‘human’), or it is the second item in a sequence ⟨ch⟩, ⟨nh⟩, ⟨lh⟩ which indicates palatalization.

The diacritics have three aims: the tilde (˜) indicates the nasal sound of [ɐ] and [o] vowels in the absence of a nasal consonant (*limão* [limẽw̄], *limões* [limõj̄]) ‘lemon(s)’; the diacritic (̄) indicates closed [ɐ], [e] and [o] vowels in words where the stress is either on the last or on the antepenultimate syllable (*pânico* [pẽniku] ‘panic’, *mês* [mɛj̄] ‘month’, *avô* [ɛvó] ‘grandfather’, *estômago* [stómɛgu] (EP) / [istómagu] (BP) ‘stomach’); the diacritic (˘) denotes the stressed vowel in words that have traditionally failed to comply with the language’s general rules (words with an antepenultimate stress such as *fábrica* [fábrikɐ] ‘factory’, *cômico* [kómiku] ‘comic’, *ética* [étikɐ] ‘ethics’, *dúvida* [dúvidɐ] ‘doubt’; words with the stress on the penultimate syllable such as *nível* [nívɛj̄] ‘level’, *órfão* [órfɛw̄] ‘orphan’; words with a last stressed syllable such as *café* [kɛfɛj̄] ‘coffee’, *avó* [ɛvõ] ‘grandmother’). The diacritic (˘) is only used in words resulting from the contraction of the preposition *a* with the definite article and pronouns beginning with *a* (*à* [a] ‘to’, *àquele* [akéli] (EP) / [akéli] (BP) ‘to that’, etc.). In Brazil, the ‘trema’ or ‘umlaut’ is also used to indicate the pronunciation of the glide [w] following a ⟨q⟩ or a ⟨g⟩ (*lingüística* [lĩgwíftikɐ] ‘linguistics’).

² In a very few cultivated words ending in ⟨en⟩, the consonant is pronounced (e.g. *abdômen* [ɛbdõmen] ‘abdomen’).

European Portuguese and Brazilian Portuguese have two notable differences in orthography: (a) consonants that are not pronounced are still written in Portugal but have been suppressed in Brazil (e.g. *direcção* (EP) / *direção* (BP) [dirɛsɐ̃w] ‘direction’, *óptico* (EP) / *ótico* (BP) [ótiku] ‘optician’); (b) words stressed on the antepenultimate syllable whenever the penultimate begins with a nasal consonant have the diacritic (˘) in EP and (^) in BP, corresponding to two different pronunciations of the vowel (e.g. *cómodo* [kómudu] / *cômodo* [kómudu] ‘comfortable’).