# Introduction

This book grew out of an overview of Portuguese designed for undergraduate seniors and graduates with varying levels of familiarity with the language and little or no training in grammar or linguistics. Some of these students were native or semi-native speakers, while others had learned the language through formal study rather than residence in a Portuguese-speaking country. As a group, they were typical of students interested in studying language as a means of communication and literary expression rather than an end in itself, and curious to learn something about the structure and use of Portuguese without going too deep into its formal analysis under any one of the linguistic theories available. Presenting the language to those students in a systematic but interesting way turned out to be a challenging and rewarding experience. Patterned on that course, this book has been organized around several key topics seen from a linguistic viewpoint rather different from the prescriptive approach usually found in grammar manuals. Owing to its general character, *Portuguese: A Linguistic Introduction* introduces only a handful of the language's main aspects, leaving out several otherwise important topics and a number of details that should be taken up at a more advanced level.

This book is divided into eight chapters, the first of which, "Portuguese in the world," surveys the external history of the language, that is the facts surrounding its development and spread beyond its original territory. The next three chapters deal with select topics of its phonology or sound system (Chapter 2, "Sounds"), its morphology or the internal organization of words (Chapter 3, "Words"), and its syntax or sentence structure (Chapter 4, "Sentences"). Aspects of the internal history of the language, that is how it grew out of popular Latin, are presented in Chapter 5, "Portuguese in time." In continuation, Chapter 6, "The expansion of European Portuguese," overviews language variation in continental Portugal, its autonomous regions of Madeira and the Azores, and Africa, Asia, and America. Chapter 7, "Brazilian Portuguese," examines in some detail the variety of the language spoken by about 80% of its speakers. Finally, Chapter 8, "Sociolinguistic issues," analyzes select topics such as diglossia, styles and registers, forms of address, profane language, and communicative strategies.

1

The premise adopted is that a language is primarily an instrument of communication that operates in a social context and is marked by a certain degree of variation. Consequently, rather than refer to an idealized version of Portuguese, this book takes into account linguistic data such as obtain in actual language use. Whereas some of the data agree with the rules found in normative grammars based on literary usage, others clash with them. Learners need to get used to this situation and accept the fact that no living language can be reduced to a set of right-or-wrong rules such as are favored by language mavens. There are issues of usage about which speakers do not always agree, and for which objective analyses cannot offer a single "correct" solution. One reason for this is that we do not possess enough data concerning those issues. Although this situation is beginning to change with the emergence of linguistic corpora, we are not yet in a position to draw definitive conclusions, even assuming such a goal is viable. If anything, corpora of real speech show that Portuguese varies considerably. This is not surprising: a language geographically so far-flung, spoken by over two hundred million people in four continents, could not fail to show a great deal of variation.

I have made an effort to make this book user-friendly by organizing the material, particularly the examples, in a legible format and cross-referencing the chapters. Thus, an indication like (2.3.2) refers to material in Chapter 2, section 3, subsection 2. Formal analysis and terminology have been kept to a minimum, but even so a certain amount of it, particularly in the chapters on phonology, morphology, and syntax, is inevitable: you cannot talk about the motor of your car without naming its parts.

I have tried to explain technical terms in the text and have added a glossary of terms (indicated by a superscript °) that students seemed to find most troublesome. Readers should find linguistic dictionaries such as Matthews 1997 or Crystal 1994, to which I am indebted, particularly useful in resolving doubts about terminology.

While regional and social variation is taken into account, unless otherwise stated the examples represent the usage of educated adult speakers and reflect the kind of language college-trained foreign learners can reasonably expect to encounter in their contacts with educated speakers of Portuguese. Individual words are glossed for their most general meaning when used to illustrate pronunciation or morphology. Consequently a word like *pena* may be glossed simply as 'feather' even though it has other meanings, such as 'pen,' 'pity,' or the third person singular of the present indicative of the verb *penar* 'to suffer.' As learners soon find out, any third person singular verb form can be used not only with the Portuguese equivalents of *he* (*ele*) and *she* (*ela*) but also with three forms meaning *you*, namely *você*, *o senhor* (m.), and *a senhora* (f.). To avoid cumbersome glosses forms such as *critica* 'he/she/you (sg.) criticizes' or *criticam* 'they/you (pl.) criticize,' the persons of the discourse are signaled by

abbreviations such as P3sg or P3pl (see Abbreviations). All glosses, rewrites of passages in non-standard Portuguese, and translations of citations are my own.

As is well known, the countries where Portuguese has official status have yet to agree on a common orthography. To be sure, some spelling differences reflect variations in pronunciation, as in Brazilian Portuguese (BP) *fato* 'fact' vs. European Portuguese (EP) *facto* 'fact' vs. *fato* 'suit of clothes' should be considered a separate word. Other differences, however, are purely visual, as in the case of the word for *linguistics*, which is written with an umlaut (*lingüística*) in BP and without it (*linguística*) in EP. There is really no good reason for such discrepancies, which are due less to language differences than to the stubbornness of those who make decisions on spelling matters, but one learns to live with them. Fortunately for learners, those spelling differences, puzzling as they can sometimes be, are really minor and do not interfere with comprehension. This book uses the current Brazilian orthography except in examples of European usage or in quoted passages and bibliographical items, which are given in their original spelling.

Such minor discrepancies are not a linguistic problem, for a language can exist with two or more partially overlapping spelling systems. Even variations in the lexicon and grammar do not necessarily compromise the integrity of the overall system, as is clearly demonstrated by the case of English, with two major international standards and other emerging regional ones (Burns and Coffin 2001; Trudgill and Hannah 2002). Whenever appropriate, I have compared the Brazilian and European varieties, which differ in various ways. For reasons discussed in Chapter 2, pronunciation can be a bit of problem for learners accustomed to BP when they are exposed to to EP, though the problem is lesser in the opposite direction. Lexical differences between BP and EP can be substantial, depending on the semantic area involved (Villar 1989, Prata 1984) but do not affect the structure of the language. Although the syntactic core is common to both varieties (as well as to the EP-related varieties spoken in Africa), there are clear contrasts in certain areas of sentence structure, such as the use (or, in the case of BP, non-use) of unstressed pronouns. There are also important contrasts in pragmatics, that is the norms of personal interaction and the strategies used to ask for something, to interrupt, to agree, to disagree, and so on. Unfortunately, little research has been carried out in the pragmatics of either variety, let alone in comparing them systematically, and consequently learners used to one variety have to feel their way around in communicating with speakers of the other. They can find solace, as well as encouragement, in the fact that native speakers face similar difficulties and usually overcome them, as demonstrated by the millions of Portuguese living in Brazil and thousands of Brazilians living in Portugal.

The often sharp differences between the educated and vernacular or popular varieties of Brazilian Portuguese can also be a source of puzzlement for

4          Portuguese: a linguistic introduction

learners. Those differences, however, are not unique and should not be blown
out of proportion: popular speech in London, New York, Mexico City, Paris,
or for that matter Lisbon can be just as impenetrable, initially at any rate, to
outsiders. An American friend of mine who had studied Portuguese for over a
year summed up this situation in a pithy message e-mailed a few weeks after
arriving in Rio: "This language is going to kill me. I'm starting to hear things on
TV and in conversation with educated people, but when it comes to the street,
it's like they speak a different language." Nevertheless he survived, traveled
around, met people, made friends, and is currently making plans to go back.
His experience, which is that of many other people, simply underscores the
fact that comprehension does not happen between languages or varieties of
a language, but between people. A speaker of variety A who is dismayed or
annoyed because speakers of variety B do not talk like the folks back home
does not have a linguistic problem but a cultural one, which can only be solved
if one is genuinely willing to work to develop the ability – which does not
come naturally – to understand the other. It is a process that can be frustrat-
ing at times, but a dash of good humor certainly helps, and a bit of linguis-
tics may be prove handy to sort out those differences and show they are not
haphazard but systematic. It is hoped that this book will be of use to serious
learners of Portuguese like my friend in coming to terms with this multifaceted
language.

# 1    The Portuguese language in the world

On March 22, 2002, in Dili, the capital city of East Timor, a national constitution was enacted, whose Article 13 stated that "*O tétum e o português são as línguas oficiais da República Democrática de Timor-Leste*," meaning 'Tetum [a Southeast Asian language] and Portuguese are the official languages of the Democratic Republic of East Timor.' A significant detail is that all the names and most of the surnames of the parliamentarians who signed the new constitution are Portuguese (*Lourdes*, *Manuel*, *Maria*, *José*, *Luisa*, *Norberto*, *Costa*, *Martins*, *Silva*, *Alves*, and so on), even though reportedly only about 2% of the population of East Timor speak Portuguese (Ethnologue.com 2002).

Having a constitution was a major accomplishment for that small country. After becoming independent in 1975 from Portugal, whose colony it had been since the fifteenth century, Timor was occupied by Indonesia for the next twenty-five years, and had to secure its freedom again at a heavy toll in human lives. Historically, however, this is just one more occasion on which Portuguese has served not only as a vital link to the outside world but also as a common language for speakers of East Timor's nineteen other languages, some of which, like Adabe or Habu, have only about one thousand speakers each. Portuguese has often played the role of a lingua franca° (a topic to be taken up again in Chapter 6) since the fifteenth century, when it began to spread from its birthplace in the Iberian Peninsula to reach the four courners of the earth.

At the time of writing (2002), Portuguese has official status in eight countries, namely Angola, Brazil (Brasil), Cape Verde (Cabo Verde), Guinea-Bissau (Guiné-Bissau), Mozambique (Moçambique), Portugal, São Tomé and Príncipe (São Tomé e Príncipe), and East Timor (Timor Leste). Spoken by about a million people in 1500, it is now estimated to be the first language of some 176 million people, a figure that shoots up to 191 million if we include secondary speakers, that is people who have learned Portuguese as a second language (www.ethnologue.com). Though approximate, such figures put Portuguese in sixth place among the languages with the largest number of speakers, after Mandarin Chinese, English, Spanish, Hindi, and Arabic (Crystal 1997:289). In what follows we will review the main points of the external history of Portuguese,

5

identify its geographic distribution to facilitate visualizing its extension on a map, and highlight other aspects of the language's situation in the world.

## 1.1 The growth of Portuguese

Portuguese is a Romance° language, like Catalan, French, Galician, Italian, Rumanian, Romansh, and Spanish (to mention only those with official status and leaving out others such as Corsican, Piedmontese, and Sardinian). Perhaps disappointingly, the word "Romance" in this context has nothing to do with Latin lovers; rather, it comes from the Latin adverb *romanice* 'in the Roman way,' which was used in the Middle Ages to designate the new speech that grew out of the popular Latin spoken in the Western Roman Empire. From the last centuries of the Empire until the emergence of Portugal as a sovereign state in the twelfth century, the language now called Portuguese gradually took shape as a spoken tongue, largely overshadowed by the universality of medieval Latin as the medium of written communication.

### 1.1.1 Roman Hispania

So far as we know, the earliest inhabitants of the Iberian Peninsula included the ancestors of the Basques in the north, the Tartesians in the south, the Iberians (a name linked to that of the Ebro River) in the east and, after about 1000 BC, the Celts, who came from northern Europe and mixed with the Iberians. Trade on the Mediterranean led to the establishment of colonies by the Phoenicians (Málaga and Cádiz), by the Carthagineans, themselves of Phoenician origin but established in Carthage, near present-day Tunis (Cartagena and Mahon, on the island of Menorca), and by the Greeks (Ampurias, on today's Costa Brava).

The Romans arrived in the Iberian Peninsula in 218 BC and colonized it in two centuries, except for a northeastern strip in today's Basque Country. They called the peninsula *Hispania* – according to a charming legend, after a Phoenician name, *i-shepham-im* or 'land of rabbits' (Eslava-Galán 1995:13) – and divided it into provinces (Map 1.1). They also provided effective administrators and troops to keep the *pax romana* 'Roman peace.' The northwestern corner of the peninsula became a province called *Gallaecia*, after its early inhabitants, the Gallaeci, a name that persists in the toponym 'Galicia' (Pg, Gal *Galiza*), which is one of contemporary Spain's autonomous regions, bounded to the east by two other autonomous regions, Asturias and Castilla-León, and to the south by the Minho River, its natural border with Portugal.

To the south and to the east of Gallaecia lay another Roman province, named *Lusitania*, after a mythical demi-god, Luso (Lat *Lusus*), a son of Bacchus and the legendary founder of Portugal. That name survives in the prefix *Luso-* 'Portuguese,' used in expressions like *estudos luso-brasileiros* 'Luso-Brazilian

Map 1.1  Schematic location of the provinces of Roman Hispania

studies' or *lusófono* 'lusophone,' that is, 'Portuguese-speaking.' It also appears in *lusofonia* 'lusophony,' a somewhat protean designation, referring loosely to the Portuguese-speaking peoples and/or regions, which was coined in the late twentieth century based on the French word *francophonie* 'francophony,' or 'French-speaking.' When the Romans defeated the Carthagenians in the second Punic War (202 BC), Lusitania was part of the Roman province of Hispania Ulterior, and its administrative center was located in what is now Mérida, in Spanish Extremadura. Nevertheless, most of Lusitania's territory from the left bank of the Minho to the Algarve region in the south makes up today's Portugal. This country's name, in turn, goes back to the Latin toponym PORTU CALE, an erstwhile Roman military installation named *Cale* which overlooked a port on the Douro river, near the site of today's city of Porto (Eng *Oporto*, so called because of the Portuguese habit of referring to "o Porto"). Over the centuries, the Latin spoken in Gallaecia would change into a Romance speech which in turn gave origin to two closely related languages, namely Galician, still spoken in that region, and Portuguese, which spread southward in the wake of conquest and settlement.

According to the Roman Empire's colonizing policy, land was provided as a bonus for retired soldiers, many of whom had served in the legions stationed in

the peninsula. Soldiers, settlers, and administrators all spoke Latin, and although
Rome did not care what language subjected peoples actually spoke – learning
Latin was considered a privilege rather than a duty – the prestige of Roman civ-
ilization, manifested by an impressive network of paved roads, bridges, aque-
ducts, temples, theaters, public baths, circuses, and an administrative organiza-
tion unparalleled in the ancient world, led the original inhabitants of Hispania
to adopt the language and customs of the Romans. After a period of bilingual-
ism, the languages spoken before the Romans' arrival – with the exception of
Basque – were eventually replaced by the settlers' popular Latin, which coex-
isted with the more cultivated variety used by officials and an educated elite.
Eventually, some cities in Hispania – Tarragona, Córdoba, Mérida – emulated
Rome in beauty and quality of life, and in the first century of our era a number of
Hispanic Romans – such as the philosophers Seneca the Elder, his son Seneca
the Younger, the poets Lucan and Martial, and the rhetorician Quintilian – made
major contributions to Latin letters.

### 1.1.2    Visigothic Hispania

The collapse of the Western Roman Empire, whose last emperor, Romulus
Augustulus, was deposed in 476 AD, brought about the breakdown of admin-
istrative cohesiveness and communication among Hispania's various regions.
The ensuing isolation led to increasing regional differentiation in the Latin
spoken in Hispania and elsewhere. Beginning in the first decade of the fifth
century, a series of invasions by Germanic tribes such as the Suevi, the Vandals,
and the Alans culminated, in the early sixth century, with the arrival of the
Visigoths. These either enslaved, killed off or drove away their predecessors,
with the exception of the Suevi, who maintained a small kingdom in Gallaecia
until about 585, when they too were conquered by the Visigoths. Being quite
romanized, thanks to an early sojourn in Roman Gaul, the Visigoths adopted
Roman Hispania's language, customs, and religion. By the end of the sixth
century the Visigothic kingdom, which had Toledo as its capital, extended all
over the peninsula, though its suzerainty over the Basque Country remained
nominal, as the Roman domination had been – a factor that permitted survival
of the Basque language until our days.

Even in the heyday of the Roman Empire the speech of the inhabitants of
the peninsula was essentially popular Latin, which differed noticeably from the
literary Latin we learn at school. By the end of the sixth century, however, that
speech had changed even further, into something considerably different from
Latin as it was still spoken and written, as a learned language, by a literate
minority associated primarily with clerical life. This new way of speaking,
whether or not people realized that it derived from Latin, became known as
*fabulare romanice*, that is 'to speak Romance.' Eventually the adverb *romanice*,
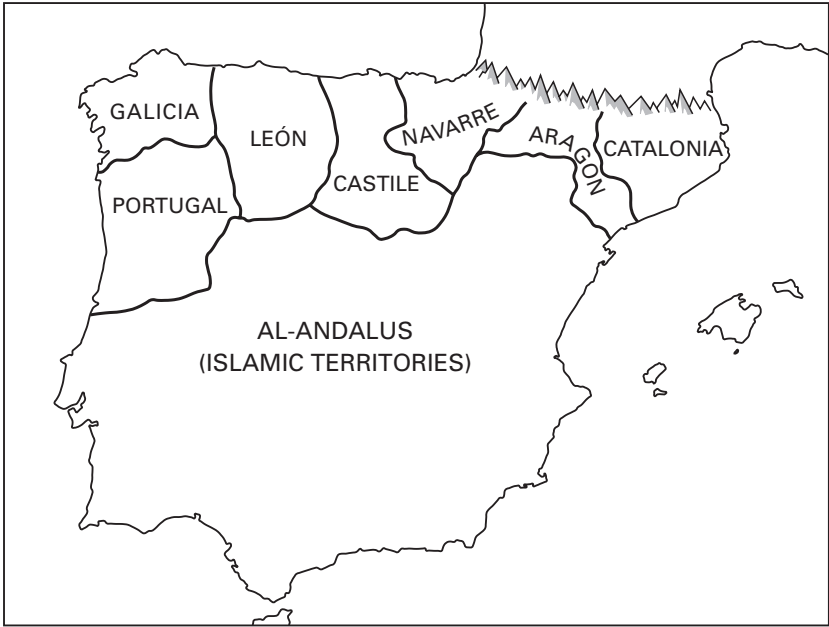
Map 1.2  Schematic location of Christian dominions and Islamic holdings in Hispania in the early tenth century

shortened to *romance*, was reinterpreted as a noun designating the local speech. In other words, everyone spoke Romance, and a few, usually members of the clergy, also learned to speak and write medieval Latin. (The verb *fabulare*, originally meaning 'to spin a yarn,' was the source of Pg *falar*, Sp *hablar* 'to talk.') Since most people were illiterate, and all writing was done in Latin, written signs reflecting a Romance speech were slow to appear, and its growth and expansion are closely linked to those of the Christian kingdoms that emerged from what had been Roman Hispania.

### 1.1.3    Islamic Hispania

In 711, the Iberian peninsula was invaded by an Islamic army made up of Berbers and Arabs, collectively called Moors° (after Lat *Mauri*, the name of the inhabitants of Mauritania in Roman North Africa). The invaders crossed the Strait of Gibraltar (from *gebel-al-Tarik* 'Tarik's hill,' a toponym that preserves the invaders' leader's name) and in a short time overwhelmed the Visigothic kingdom. By 718 the peninsula had been divided (Map 1.2) into a southern Islamic area, Al-Andalus (a name that recalls the ferocious Vandals and is preserved today in the place-name Andalusia), and a northern fringe under

Christian domination. The latter area, which included a few strongholds in the
mountains of Asturias, the Basque region, and a string of fortifications called the
Spanish March, set up by Charlemagne (742–814) along the Pyrenees, would
eventually be divided into several Christian kingdoms and counties which, for
the next seven centuries, fought to reconquer the territory lost to the invaders.
Also in 718, the Christians holding out in the mountains of Asturias achieved a
small victory over a detachment of Moors in a skirmish celebrated in legend as
the battle of Covadonga, traditionally held to be the beginning of the reconquest
which culminated in the fall of Granada to the Catholic Monarchs Ferdinand
and Isabella in 1492.

## 1.2      The formation and expansion of Portuguese

A fundamental linguistic consequence of the Arab conquest of the Iberian Penin-
sula was that in each of the small northern territories the Romance vernacular°
developed features that differentiated it from the speech of neighboring regions.
The speech of Christians in Al-Andalus itself, known as Mozarabic° (from Ar
*musta'rib* 'Arab-like'), was eventually absorbed by the northern Romances as
these expanded into territory conquered from Islam. In the Christian area, start-
ing at the Mediterranean end of the Pyrenees, in the region around Barcelona,
there arose Catalan, which was carried to Valencia, Alicante, and the Balearic
Islands in the first half or the thirteenth century by military conquest. Immedi-
ately to the west a group of closely related vernaculars known to linguists as
Navarro-Aragonese developed. In an initially small area around and north of
Burgos, there was another Romance known as Castilian, which in time would
extend over most of the peninsula and develop into modern Spanish. To the
west of Castilian there developed a Romance speech known as Leonese, or
Asturian-Leonese, parent of the various *bables*, as the vernaculars spoken in
Asturias are still called. Finally, in the northwest corner of the peninsula, in the
former Roman province of Gallaecia, was born the vernacular which linguists
refer to as Galician-Portuguese, the parent of modern Galician as well as of
Portuguese.

By the eleventh century the kingdom of Castile (Sp *Castilla*) had acquired
hegemony over Leon and Galicia and was leading the reconquest of Muslim
Spain. Its policy of establishing alliances with the Franks beyond the Pyre-
nees entailed an infusion of French culture, such as the creation of monasteries
linked to Cluny Abbey in Burgundy, and the substitution of Roman liturgy
for the ancient Hispanic (often called "Mozarabic") liturgy inherited from the
Visigothic kingdom. Another consequence was the presence of French noble-
men who came to seek fortune in frequent campaigns waged against the Muslim
states of Al-Andalus.

One of these French adventurers, Count Raymond of Burgundy, a region
in central eastern France, in 1087 married Urraca (a charming name meaning