Cambridge Language Surveys

The Slavic Languages

Roland Sussex and
Paul Cubberley

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521223157

This page intentionally left blank

THE SLAVIC LANGUAGES

The Slavic group of languages – the fourth largest Indo-European subgroup – is one of the major language families of the modern world. With 297 million speakers, Slavic comprises 13 languages split into three groups: South Slavic, which includes Bosnian, Serbian and Croatian; East Slavic, which includes Russian and Ukrainian; and West Slavic, which includes Polish, Czech and Slovak. This book, written by two leading scholars in Slavic linguistics, presents a survey of all aspects of the linguistic structure of the Slavic languages, considering in particular those languages that enjoy official status. As well as covering the central issues of phonology, morphology, syntax, word-formation, lexicology and typology, the authors discuss Slavic dialects, sociolinguistic issues and the socio-historical evolution of the Slavic languages.

Accessibly written and comprehensive in its coverage, this book will be welcomed by scholars and students of Slavic languages, as well as by linguists across the many branches of the discipline.

ROLAND SUSSEX is Professor of Applied Language Studies at the University of Queensland, and formerly Professor of Russian at the University of Melbourne. He has taught a wide variety of courses in linguistics and applied language studies, including the linguistic description of the Slavic languages. He has previously published *A Bibliography of Computer-Aided Language Learning* (with David Bradley and Graham Scott, 1986), and *Computers, Language Learning and Language Teaching* (with Khurshid Ahmad, Margaret Rogers and Greville Corbett, Cambridge University Press, 1985).

PAUL CUBBERLEY was Senior Research Fellow in Linguistics at the University of Melbourne until 2001, and was previously Head of Russian there. He has also taught Czech, Polish, Old Church Slavonic, comparative Slavonic linguistics and the history of the Russian language. His previous publications include *The Suprasegmental Features in Slavonic Phonetic Typology* (1980), and most recently *Russian: a Linguistic Introduction* (Cambridge University Press, 2002).

CAMBRIDGE LANGUAGE SURVEYS

General editors
P. Austin (*University of London*)
J. Bresnan (*Stanford University*)
B. Comrie (*Max Planck Institute for Evolutionary Anthropology, Leipzig*)
S. Crain (*University of Maryland*)
W. Dressler (*University of Vienna*)
C. Ewen (*University of Leiden*)
R. Lass (*University of Cape Town*)
D. Lightfoot (*University of Maryland*)
K. Rice (*University of Toronto*)
I. Roberts (*University of Cambridge*)
S. Romaine (*University of Oxford*)
N. V. Smith (*University College, London*)

This series offers general accounts of the major language families of the world, with volumes organized either on a purely genetic basis or on a geographical basis, whichever yields the most convenient and intelligible grouping in each case. Each volume compares and contrasts the typological features of the languages it deals with. It also treats the relevant genetic relationships, historical development and sociolinguistic issues arising from their role and use in the world today. The books are intended for linguists from undergraduate level upwards, but no special knowledge of the languages under consideration is assumed. Volumes such as those on Australia and the Amazon Basin are also of wider relevance, as the future of the languages and their speakers raises important social and political issues.

Volumes already published include
Chinese *Jerry Norman*
The languages of Japan *Masayoshi Shibatani*
Pidgins and Creoles (Volume I: Theory and structure; Volume II: Reference survey) *John A. Holm*
The Indo-Aryan languages *Colin Masica*
The Celtic languages *edited by Donald MacAulay*
The Romance languages *Rebecca Posner*
The Amazonian languages *edited by R. M. W. Dixon and Alexandra Y. Aikhenvald*
The languages of Native North America *Marianne Mithun*
The Korean language *Ho-Him Sohn*
Australian languages *R. M. W. Dixon*
The Dravidian languages *Bhadriraju Krishnamurti*
The languages of the Andes *Willem Adelaar with Pieter Muysken*
The Slavic languages *Roland Sussex and Paul Cubberley*

THE SLAVIC LANGUAGES

ROLAND SUSSEX

PAUL CUBBERLEY

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521223157

© Roland Sussex and Paul Cubberley 2006

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2006

ISBN-13 978-0-511-23947-2 eBook (Adobe Reader)
ISBN-10 0-511-23947-5 eBook (Adobe Reader)

ISBN-13 978-0-521-22315-7 hardback
ISBN-10 0-521-22315-6 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

For Bogusia and Gladys
Matthew and Joanna
Nadine and Michelle

CONTENTS

PREFACE

Like Rebecca Posner, whose *The Romance languages* in this series was published in 1996, we have often been daunted by the size and complexity of the task. Slavic is not only a large group of languages but it is also the most written-about (see the Introduction).

Over the last decade Slavic has also been arguably the most externally unstable of the Indo-European language families. The fall of Euro-Communism was marked most dramatically by the tearing down of the Berlin Wall in 1989 and the subsequent dissolution of the USSR, the Warsaw Pact, COMECON and the structures and infrastructures of what US President Ronald Reagan called the "evil empire". What emerged from the political rebirth of the Slavic lands has turned out to be a linguistic landscape where some stable and persisting features have found themselves side-by-side with a dynamic, unstable and volatile cultural context. The Slavic languages are living in more than just interesting times. It has been our task to try and seize them in motion. We have spent most of our academic lives working in these languages, and this book is partly by way of thanks to the stimulus that working in and on Slavic has given us, and to our colleagues in Slavic around the world who have contributed to the discipline.

A book like this has a dual audience. On the one hand we are addressing Slavists who need to widen their knowledge about other Slavic languages – scholars who know some Russian, say, and are curious about the other Slavic languages. On the other hand are students and scholars of languages and linguistics with no particular knowledge of a Slavic language. This double focus makes for difficulties of selection and presentation. We have tried to write to, and for, both audiences.

We have also tried to meet the needs of both the consecutive reader and the reference reader, who needs to find how, say, questions are formed in East Slavic. We do tell a story of Slavic, but the text and index are structured so as to make it possible to locate specific issues, and cross-references allow navigation through such issues.

The survey is a relentless genre. In a hugely documented language family like Slavic, survey authors are constantly faced with major decisions of omission, inclusion and angle of view. We have had to weigh our favorite crannies of a language, and cherished idiosyncrasies of this corner of phonology, or that lexical

oddity, against the big picture, but at the same time a big picture with enough detail to give a true feel for what snaps into focus when the camera zooms in, as well as a valid panorama when it zooms out again.

We have culled the examples for what we analyze as the eleven Slavic languages from many sources, published, web-based, oral and personal, and we are sometimes not sure where some of them originated. We apologize to previous authors if we have borrowed their examples without acknowledgment, and invite colleagues to use ours freely in the same spirit of scholarly investigation.

ACKNOWLEDGMENTS

Among the many friends and colleagues who have contributed to the preparation of this book, we would like to extend particular thanks to Ian Press, who read the manuscript for the Press and made many invaluable suggestions; to Bernard Comrie, Grev Corbett, Roger Lass, John Lyons and Ron Sussex; to Judith Ayling, Kate Brett and Helen Barton at Cambridge University Press; and, most of all to our families, who have patiently supported this complex project over many reefs and through many mountain passes.


Czech
Slovak
Slovenian
Bosnian/
Serbian/
Croatian
Macedonian
Bulgarian
Russian
Ukrainian
Belarusian
Polish
Sorbian
Upper
Lower
NORWAY
FINLAND
NORWAY
SWEDEN
RUSSIA
DENMARK
BELARUS
Lower
Upper
POLAND
GERMANY
UKRAINE
CZECH
REP.
SLOVAKIA
AUSTRIA
HUNGARY
ROMANIA
ITALY
BLACK SEA
BULGARIA
ALBANIA
TURKEY
GREECE

0

Introduction

0.1 Survey

This book presents a survey of the modern Slavic languages – known as "Slavonic" languages in Britain and some of the Commonwealth countries[1] – seen from the point of view of their genetic and typological properties, their emergence and standing as national languages and selected sociolinguistic characteristics.

The language survey as a genre, and as defined in the description of this series, is not the same as a comprehensive comparative grammar. The survey does require breadth, to cover the full range of languages; and selective depth, to identify and highlight the specific properties of the language family as a whole, and the properties of sub-families and languages within the family. Our treatment is deliberately selective, and we concentrate on topics and features which contribute to the typology of the members of the Slavic language family.

We have tried to achieve this balance with two goals in view: to present an overview of the Slavic languages, combined with sufficient detail and examples to form a sound empirical basis; and to provide an entry point into the field for linguistically informed and interested readers who do not already command a Slavic language.

0.2 The Slavic languages in the world

The Slavic languages are one of the major language families of the modern world. In the current world population of over 6 billion the most populous language family is Indo-European, with over 40 percent. Within Indo-European Slavic

[1] North America favors "Slavic", while (British) Commonwealth countries prefer "Slavonic". North Americans usually pronounce "Slavic" and "Slavist" with the vowel [a], corresponding to [æ] in British English. This dual nomenclature is not found in other major languages of scholarship: French *slave*, German *slawisch*, Russian *slavjánskij*.

is the fourth largest sub-family, with around 300 million speakers, after Indic, Romance and Germanic, and ahead of Iranian, Greek, Albanian and Baltic.

0.3 **Languages, variants and nomenclature**

Modern Slavic falls into three major groups, according to linguistic and historical factors (table 0.1). We shall concentrate on the Slavic languages which enjoy official status in modern times, and have an accepted cultural and functional standing: Slovenian; Croatian, Bosnian and Serbian; Bulgarian and Macedonian in South Slavic (2.2); Russian, Belarusian and Ukrainian in East Slavic (2.3); and Upper and Lower Sorbian, Polish, Czech and Slovak, in West Slavic (2.4). The status of Bosnian, Croatian and Serbian is explained in more detail below. In listing Montenegrin as a "sub-national variety" we simply assert that it is not the language of an independent state, nor an officially designated language.

We shall also make considerable reference to Proto-Slavic, the Slavic dialect which emerged from Indo-European as the parent language of Slavic; and to Old Church Slavonic, originally a South Slavic liturgical and literary language now extinct except in church use, which is of major importance as a cultural, linguistic and sociolinguistic model. Both Proto-Slavic and Old Church Slavonic are fundamental to an understanding of the modern languages, especially in the chapters

Table 0.1. *Modern Slavic families and sub-families*

	National languages	Sub-national varieties	Extinct languages
South Slavic	Slovenian		
	Croatian	Montenegrin	
	Bosnian		
	Serbian		
	Bulgarian		Old Church Slavonic
	Macedonian		
East Slavic	Russian		
	Belarusian	Rusyn (Rusnak)	
	Ukrainian	Ruthenian	
West Slavic	Sorbian (Upper and Lower)		
	Polish	Kashubian	Polabian
			Slovincian
	Czech	Lachian	
	Slovak		

on phonology and morphology. Other Slavic languages/dialects will be used as relevant for illustration and contrast.

The modern Slavic languages exhibit a moderate degree of mutual comprehensibility, at least at the conversational level. The ability of Slavs to communicate with other Slavs across language boundaries is closely related to linguistic and geographical distance. East Slavs can communicate with each other quite well. So can Czechs and Slovaks, Poles and Sorbs, and indeed all West Slavs to some extent. Among the South Slavs, Bulgarian and Macedonian are inter-communicable, as are Bosnian, Croatian and Serbian, to varying degrees (2.2.4).

The Slavic languages and variants discussed in this book are listed in table 0.1. We adopt the convention of listing the three major families in the order South, East, West, which allows for a convenient discussion of historical events. Within each family we follow the order north to south, and within that west to east; languages in columns 2 and 3 are related to those within the same sub-family in column 1. The geographical distribution of the national languages is shown in the map on page xx.

There are also some important issues of nomenclature. The names of the languages and countries in English can vary according to convention and, to some extent, according to personal preference. We use the most neutral current terms in English. A useful distinction is sometimes made in English between the nominal ethnonym and the general adjective, e.g. "Serb", "Slovene", "Croat", for the ethnonym *vs* "-ian" for the adjective: "Serbian", "Slovenian", "Croatian"; "Slav" is also used as an ethnonym. We have used "-ian" for the languages, following common practice.

The word "language" has a major symbolic significance among the Slavs. A variety which warrants the label "language" powerfully reinforces the ethnic sense of identity. Conversely, "variants" look sub-national and so lack status and prestige. A typical case is Croatian: under Tito's Republic of Yugoslavia, Croatian was one of the two national variants of Serbo-Croatian. But the Croats fought vigorously from the 1960s for the recognition of Croatian as a "language", for instance in the constitution of the Republic of Croatia (Naylor, 1980), a battle which they won with the establishment of the independent Republic of Croatia in 1991.

The criteria relevant to language-hood also vary. For any two variants, the factors which will tend to class them as languages include mutual unintelligibility, formal differentiation, separate ethnic identity and separate political status. Sometimes politics and ethnicity win over intelligibility, as happened with Croatian and Serbian, and now with the recently created Bosnian: Bosnia entered the United Nations in 1992, accompanied by the emergence of the Bosnian language. Sorbian presents a very different profile: numerically small in population

terms, and with no political autonomy, Upper and Lower Sorbian show significant formal differences, though they are mutually intelligible to a substantial degree. We have classed them as variants of a single language, Sorbian. The reasons for such classifications for different languages and varieties are given in chapter 3. We aim broadly, where the linguistic data warrant it, to respect the declared identity and linguistic allegiance of the different Slavic speakers. In using the term "language" we mean a defined variety with formal coherence and standardization, and some cultural and political status.

0.3.1 South Slavic

"Yugoslavia" is also written "Jugoslavia", or "Jugoslavija", following Croatian usage. The name means "south Slavdom". We favor "Yugoslavia" as being more common in English usage. Until 2003 Serbia, including Montenegro, continued to use the name Yugoslavia/Jugoslavia. The Federal Republic of Yugoslavia (commonly abbreviated as "FRY") was admitted to the United Nations in 2000. In early 2003 legislation paved the way for the separation of Montenegro and Serbia within three years.

"Slovenian" is also known as "Slovene", especially in British usage. We prefer the former, bringing it into line with the other South Slavic languages Bosnian, Croatian (also "Croat"), Serbian, Macedonian and Bulgarian.

"Croatian", "Bosnian" and "Serbian" merit special comment, and the relation of Croatian and Serbian to each other and to "Serbo-Croatian" (also "Serbo-Croat") is culturally, ethnically and linguistically highly sensitive (2.2.4). Serbo-Croatian was negotiated in 1850 as a supra-ethnic national language to link the Serbs and Croats. It survived with some rough periods until the 1980s, when it was effectively dissolved by the secession of the Croats as they established an independent Croatia. Bosnia then separated from Serbia in 1992. "Serbo-Croatian" is consequently now an anachronism from the political point of view, but there is still an important linguistic sense in which Croatian, Bosnian and Serbian belong to a common language grouping. For this reason we use the abbreviation "B/C/S" to cover phenomena which are common to these three languages. We shall use "Serbo-Croatian" in relation to scholarship specifically referring to it (or to common elements of the former standard, now the three modern standards). "Bosnian" was generally assumed to be included under "Serbo-Croatian" before the creation of the state of Bosnia. Bosnian, Croatian and Serbian will be relevant to such scholarship in differing degrees.

Montenegrin, a western variety of Serbian, has also been proposed for language-hood by Montenegrin nationalists. However, Montenegrin is not fully standardized,

and is properly considered at this stage as a sub-national western variety of Serbian.

"Bulgarian", though the name originally belonged to a non-Slavic invader (2.2.2), is an uncontroversial name for the language and inhabitants of contemporary Bulgaria, reinforced by more than a millennium of literacy.

The new name of the former Yugoslav Republic of Macedonia (also "FYROM") is currently unresolved, with the Greek government claiming prior historical rights to the name "Macedonia". In this book we shall use "FYR Macedonia", the common current political compromise for the "Former Yugoslav Republic of Macedonia". The name of the Macedonian language is also disputed by the Greeks and Bulgarians, but we shall follow the established Slavists' convention and use "Macedonian", since there is no other language competing for this name (2.2.3).

0.3.2 East Slavic

"Russian" was also known as "Great Russian", a term dating from the days of Imperial Russia (–1917), though it was also used in the Russian-language imperialist policies of the USSR, especially in the 1930s and 1940s. "Great Russian" is now of historical interest only. "Russia" was sometimes loosely used in English during the time of the USSR to refer to the USSR itself.

"Ukrainian" was formerly known as "Little Russian", as distinct from "Great Russian", and thought by some to imply that Ukrainian was a subordinate variety. This term has now been erased by nationalistic pressures in Ukraine. Ukrainian has sometimes also been misnamed "Ruthenian", a name used especially before 1945, when much of this area became part of Soviet Ukraine, to designate the Transcarpathian dialects around Prešov in Slovakia. Nowadays "Ruthenian" is used mainly for immigrants from this area in the USA and in the Vojvodina area of former Yugoslavia (Shevelov, 1993: 996).

Ukrainians prefer English "Ukraine" to "The Ukraine" and Russian "*v Ukraíne*" to "*na Ukraíne*" for 'in Ukraine'. In each case the second form suggests a region rather than a country. We shall follow their English preference.

"Belarusian" was known as "Belorussian" before independence in 1991, reflecting the Russian spelling of the language. Belarusians have always used (and still use) *belarúskij*. After the dissolution of the USSR, national sentiment moved the Belarusians to differentiate their language from Russian. Belarusian was also known in English as "White Russian" (the root *bel-* means 'white'). The official name of the modern country is *Belarus* (Belarusian *Belarús'*), not the former *Belorussia*, hence "Belarusian" is the most suitable English form; some also call it

"Belarusan". These names have no specific connection with the anti-Communist White Russians of the years following the Russian Revolution.

The name "Rusyn" (or "Rusnak") has been used in various senses, sometimes overlapping with Ukrainian. There is disagreement over whether Rusyn is a dialect of Ukrainian or independent. One contemporary designation is for a group of about 25,000–50,000 speakers of an East Slovak dialect who now live in the Vojvodina area of Yugoslavia. Magocsi (1992) marks the proclamation of a new Slavic literary language in East Slovakia, the west of Ukraine and south-east Poland around Lemko. They claim 800,000–1,000,000 speakers for Rusyn. This declaration has not so far been matched by wider recognition outside the Rusyn area. Shevelov calls Rusyn 'an independent standard micro-language' (1993: 996).

0.3.3 West Slavic

"Czechoslovak", sometimes used for the language of the former Czechoslovak Republic, is a misnomer. Czech and Slovak are distinct languages, and the official languages of the modern Czech and Slovak Republics, respectively.

"Sorb" and "Sorbian" are equivalent, but this language is also sometimes known as "Wendish", a term which now can have pejorative connotations in German, and which must also be distinguished from "Windish" or "Windisch", the name normally given to a group of Slovenian dialects in Austria. "Lusatian", another name used for the language (e.g. by de Bray, 1980c), properly refers to any inhabitant of Lusatia, the homeland of the Sorbs in modern Germany, irrespective of race or language. The language is also sometimes known as "Saxon Lusatian" and "Sorabe" (the regular French term). We shall use "Sorbian", following Stone, (1972, 1993a), leaving "Sorb" as the ethnonym.

Polish is one of the least controversial ethnonyms and linguanyms among the Slavs. Although the political status of Poland has varied widely across the centuries, the area of the Poles and the Polish language, centred approximately around Warsaw, have been relatively more stable.

Rusyn, which overlaps between the West and East Slavic areas, is discussed above under "East Slavic".

Kashubian (Polish *kaszubski*) is also known as "Cassubian" (Stone, 1993b). Although it is variously reported as numbering around 300,000 speakers, Ethnologue (see table 0.2) has it at only 3,000, with most of the speakers using dialectal Polish. Kashubian lacks most of the linguistic and social determinants of language-hood, and we will treat it as a north-western variety of Polish.

In this book we regard Lachian, a numerically small variety of Czech, as a dialect.

Table 0.2. *Slavic languages: numbers of speakers following Ethnologue (www.sil.org)*

Language	Total speakers	Homeland speakers	Country
South Slavic			
Serbo-Croatian	21 million	10.2 million	Bosnia, Croatia, Serbia
Bulgarian	9 million	8 million	Bulgaria
Slovenian	2 million	1.7 million	Slovenia
Macedonian	2 million	1.4 million	FYR Macedonia
East Slavic			
Russian	167 million	153.7 million	Russia
Ukrainian	47 million	31.1 million	Ukraine
Belarusian	10.2 million	7.9 million	Belarus
West Slavic			
Polish	44 million	36.6 million	Poland
Czech	12 million	10 million	Czech Republic
Slovak	5.6 million	4.9 million	Slovak Republic
Kashubian	3,000	3,000	Poland
Sorbian	69,000	69,000	Germany
Total (millions)	319.8 +	265.5 +	

Note: The figures for Sorbian are Ethnologue's estimate for total speakers, but many are Sorbian ~ German bilinguals, and the total figure for primary users is probably under 30,000.

A common feature of the modern Slavic "literary" languages – the Slavs use this term for the written standard – is a strong regard for the integrity of the national language as a kind of symbolic monument. There is strong centralized regulation of the language, and highly developed "corpus planning" to establish and maintain the languages' identity and purity. While regional variation is acknowledged and encouraged, social variation and sub-"literary" use are treated with some caution. This care for managing the languages is one of the strongest continuities between pre-Communist and Communist conceptions of language. It has begun to break down in the post-Communist era, when the concept of "literary" language has been broadened to allow much more slang, vernacular usage and creativity, including borrowing from Western languages, especially English.

Since the fall of Communism there has also been a growing pressure to ethnic self-determination, which has resulted in the emergence of national language movements in several areas of the Slavic world. We shall discuss the re-differentiation of Belarusian and Ukrainian from Russian in chapters 2 and 11. Most of the tension in this area has been between Russian and the non-Slavic members of the Confederation of Independent States. For South Slavic, the main issue is that

of B/C/S, but Montenegrin is another potential candidate for language-hood. All these tensions involve a complex mixture of political autonomy and language politics.

0.4 **Languages, polities and speakers**

The territorial adjustments following the Second World War made the political boundaries of the modern Slavic nations coincide to a larger extent than before with major linguistic and ethnic boundaries, though they were still far from being a perfect match, as can be seen from the sad violence in Yugoslavia over the past decade.

After the fall of Euro-Communism in 1989–1991 there were additional geopolitical changes. The Czechs and Slovaks separated smoothly into the Czech Republic and the Slovak Republic in the "Velvet Divorce" of 1992. Elsewhere the changes were more violent. The USSR fell apart. Ukraine and Belarus emerged as sovereign states, and as Russia lost its dominant position numerous states became autonomous: the Baltic states of Latvia, Lithuania and Estonia, the Caucasus states of Armenia and Georgia, and Central Asian states like Uzbekistan and Turkmenistan. This struggle is continuing along ethnic, political, linguistic and religious (Islamic/non-Islamic) lines, most notably in Chechnya.

The situation was especially unstable in Yugoslavia, where the federation had been held together mainly by Tito's ability as president. After his death in 1980 the components soon separated along ethnic-linguistic lines into Slovenia, Croatia, Serbia (incorporating Montenegro, Vojvodina and Kosovo, the last under UN protection since 1999), Bosnia (independent from 1992) and FYR Macedonia. The supra-ethnic national language Serbo-Croatian came to an effective end as a political and national force.

There is also a substantial Slavic diaspora. Most of the Slavic languages are spoken by significant émigré groups, especially in North America, Western Europe and Australasia, which places different geographical and cultural pressures on the languages (chapter 11). However, migration from the Slavic lands for ideological and economic reasons has now slowed, and a number of former émigré refugees are now returning home.

Table 0.2 summarizes the total numbers of homeland speakers, and the totals including speakers in émigré communities. The data for émigré speakers are not wholly reliable or comparable, since censuses in different countries have counted language ability and identity in different ways. And the total figures for homeland speakers may include minority ethnic groups: for Russian, for instance, the figure of 153.7 million includes about 16 million ethnic non-Russians. It is also important

to remember that the status of Russian was considerably enhanced by its widespread use as a second "native" language in the former USSR, and as a major foreign language of education, administration, defence, culture and trade in the other Slavic and East European countries before the disintegration of the Soviet Bloc (see chapter 11). Including speakers of Russian as a second language, the total figure for Russian was 270 million in the USSR. Since 1990 the second-language status of Russian is being replaced by other languages, especially English.

0.5 Genetic classification and typology

Slavic provides us with many examples like (1), a characteristic instance of genetic differentiation. Here the East Slavic languages have an extra syllable, a phenomenon known as "pleophony", Russian *polnoglásie* (for transcription conventions and diacritics, see below and appendix B):

(1) 'milk', 'road'

South Slavic		West Slavic		East Slavic	
Slovenian:	*mléko*	Russian:	*molokó*	Sorbian:	*mloko*
	drága		*doróga*		*droga*
B/C/S:	*mléko*	Belarusian:	*malakó*	Polish:	*mleko*
	drȁga		*daróha*		*droga*
Macedonian:	*mleko*	Ukrainian:	*molokó*	Czech:	*mléko*
			doróha		*dráha*
Bulgarian:	*mljáko*			Slovak:	*mlieko*
					dráha

There are other phenomena which distinguish West Slavic from East and South Slavic, or South Slavic from East and West Slavic, or Slavic sub-families from each other. In chapters 1 and 3–9 we shall show historically based patterns which support: a West *vs* East + South divide; a South *vs* West + East divide; a North *vs* South divide (Polish and Sorbian + East Slavic *vs* Czech, Slovak + South Slavic); and sub-patterns within the three major Slavic sub-families, with Polish and Sorbian contrasted to Czech and Slovak in West Slavic; and Bosnian, Croatian and Serbian contrasted to Bulgarian and Macedonian in South Slavic. Some features cut across internal Slavonic boundaries. The fixed-stress languages, for instance, are West Slavic and Macedonian. And there are religiously based typological features, as when Orthodox countries write in the Cyrillic script (appendix B) and tend to favour Greek-based lexis; while the non-Orthodox write in Roman and often show a greater preference for indigenous or Western lexis (chapter 9).

Moreover, not all the properties of the languages form such intra-Slavic genetic groupings. Balkan languages like Bulgarian and Macedonian (Slavic), Romanian (Romance) and Albanian (an Indo-European language-isolate), in a geographically coherent area, are genetically part of different sub-branches of Indo-European, but show similar post-posed article forms. These typological features cut across the underlying genetic classification, and since Trubetzkoy they have been covered by the term *Sprachbund* 'language union'. Balkan languages are one of the standard examples:

(2)

	'city'	'the city'
Bulgarian:	*grad*	*gradắt*
Romanian:	*oraş*	*oraşul*
Albanian:	*qytét*	*qytéti*

Macedonian shares this feature with Bulgarian, and they stand apart from the rest of South Slavic and other standard Slavic languages, though there are some dialects, for instance in Russia, where post-posed articles are also found. They also occur in Scandinavian languages.

Other *Sprachbund* features are not far to seek. Greek, Romanian, Albanian, Bulgarian and Macedonian, for instance, all lack an infinitive, and express 'I want to go' as 'I want that I may go' (Joseph, 1983). And the imperfect and aorist tenses have now effectively disappeared from all the Slavic languages except two in South Slavic (Bulgarian and Macedonian) and two in West Slavic (Upper and Lower Sorbian: 5.5.5.4). In Serbian and Croatian they are archaic or restricted to formal/literary style.

Slavic, then, exhibits both strong internal cohesion and some distinctive features which link it to adjacent language families and groups within Indo-European, especially Baltic (chapter 1). These properties form many intersecting groupings, which can lead to highly complex and articulated typologies. Our goal is not a formal typology or taxonomy in the sense of Greenberg (1978), but rather a broader and less formal, as well as less theoretically driven, treatment. This approach will be closely linked to the requirements of a linguistic survey.

0.6 The linguistics of Slavic: empirical and theoretical characteristics

The systematic comparative study of Slavic, in the opinion of many Slavists, dates from the *Vergleichende Grammatik der slavischen Sprachen* (1875–1883) by Miklosich (1825–1874), which has been to Slavic what Grimm has been to Germanic. A great deal of the comparative work on Slavic since Miklosich has followed his example in concentrating on the diachronic study of the languages.

The bulk of this work until the last few decades has been written in the Slavic languages, and in German and French.

The only four extended studies of Slavic available in English are Entwistle and Morison's *Russian and the Slavonic languages* (1964), a diachronic analysis with a bias towards Russian data, which is now dated; de Bray's *Guide to the Slavonic languages* (1980a–1980c), which is synchronic but not contrastive, since it treats each language separately; Horálek's *An introduction to the study of the Slavonic languages* (Horálek, 1992), a translation and amending by Peter Herrity of the original *Úvod do studia slovanských jazyků*, first published in Czech in 1955 and with a strong historical bias; and *The Slavonic languages*, edited by Comrie and Corbett (1993), a more structured, synchronic and theoretically oriented treatment on a language-by-language basis. Two recent cross-Slavic studies are by Dalewska-Greń (1997), a comparative study mainly concerned with phonology, morphophonology, morphology and syntax; and Panzer (1991a), which combines single-language summaries with a short treatment of Slavic typology and a historical treatment of phonology and morphology. The present book differs from its predecessors in approaching Slavic from both a synchronic and a contrastive-typological point of view, following the model of the other books in this series. We also incorporate a deliberate diachronic perspective through much of the book, in the spirit of Townsend and Janda (1996; and see the end of this section).

Modern Slavic linguistics has now happily moved beyond the interference of political ideology which was typical of much of the twentieth century (L'Hermitte, 1987), and the artificial opposition of "Western" and "non-Western" models. The Russian Revolution forced the emigration of Slavists like Trubetzkoy and Jakobson, to the advantage of the Prague Linguistic Circle. After the Second World War Jakobson's further relocation to the USA helped to establish the strong North American tradition of Slavic linguistics. Under Communism in the Soviet Union from the late 1920s into the 1950s linguistics in general, and Slavic linguistics in particular, were hindered by an anti-Western and often xenophobic ideology which eschewed "empty formalism" (i.e. structuralism and "Western", "bourgeois" models, including much of diachronic linguistics). The thaw *vis-à-vis* Western linguistics dates from 20 June 1950, when an article under Stalin's name in *Pravda* opened the way for the judicious incorporation of Western linguistics, including philological and structuralist frameworks and methodologies. But for much of the century Slavic linguistics was a matter of the West and the rest. In the West there were continuations of the powerful Moscow and Prague traditions; Jakobson's school at Harvard and the dominant effect of his students; and theory-based approaches deriving from current advances, especially in generative and later functional grammar. The reintegration of non-Western and Western Slavic

linguistics – though this happened more rapidly in Yugoslavia, Poland, the then Czechoslovakia and the DDR – took much of the rest of the century.

Soviet and East European Slavic linguistics under Communism made solid contributions to the empirical collection, analysis and description of the languages, particularly national Slavic languages, and the establishment of authoritative norms and reference materials. So too has research into the history of the national languages, albeit often from a single-sided historical-ideological perspective. This has been particularly important for more recent literary languages like Macedonian, which lacked both historical validation and contemporary authentication in dictionaries, grammars and a stable written language. Dialectology has been strongly supported, with publications in the form of dialect atlases, partly in the context of the drive to reaffirm the national languages and their regional varieties. The same cannot be said for most of the rest of sociolinguistics, since it was difficult to synchronize socially determined language variation with an ideology which promoted a classless society. The first Soviet books on sociolinguistics were published only in the late seventies (Nikol′skij, 1976), though the Prague School tradition of integrated linguistics, culture and sociolinguistics fared rather better. Lexicology and lexicography, however, have been more intensively studied in both the standard languages and regional variants than in the West, where these areas rather languished in the face of the dominance of syntax between 1957 and the 1980s. In the core of Slavic linguistics, however, there was a backing-away from structural descriptive methods, particularly those which evoked models currently fashionable in the West. Outside the Soviet Union the hand of Moscow was less oppressive, at least from the late 1950s, which helped – for instance – the continuation of the existing traditions in research on historical linguistics and Indo-European in the work of linguists like Bednarczuk (1968–1988) and Kuryłowicz (1964) in Poland, Gamkrelidze (Gamkrelidze and Ivanov, 1991) in Georgia, and Georgiev (1981) in Bulgaria. There has also been the controversial Nostratic theory, which attempts to prove that Indo-European belongs to a more widely based proto-language (Illič-Svityč, 1971–1976, 1979).

In order to strike a reasonable balance between Soviet, post-Soviet and Western sources, the theoretical orientation of this book will be as neutral as possible within the major trends in each of the linguistic disciplines. In phonology, morphophonology and morphology we have tended to follow the approach of Comrie and Corbett (1993) (chapters 3–5). The syntactic approach will also be fairly theory-neutral. In the former Soviet Union this discipline has been influenced by the Russian classification of construction-types known as *slovosočetanija* ('word-combinations'), which has been taxonomic rather than theoretical in orientation, and by the Prague School's conception of Functional Sentence Perspective (7.5).

A characteristic feature of modern Slavic linguistics is the extent to which the field, both in this book and more generally, is closely tied to the historical study of the languages. This is no accident. Nineteenth-century Slavic linguistics, like Germanic and Romance linguistics of the period, had a strongly diachronic bias. What distinguishes Slavic is the continuation of the diachronic tradition into modern mainstream Slavic linguistics. In Germanic and Romance linguistics there was a reorientation to a more synchronic emphasis in the years following the publication of de Saussure's *Cours de linguistique générale* (1916). Matters never went so far in Slavic linguistics. The new interest in synchrony, which had been anticipated about twenty years before Saussure by Baudouin de Courtenay, a Polish linguist who spent much of his working life in Russia, was well established in major centers of linguistic research in Kazan, Moscow and St. Petersburg (Stankiewicz, 1972). And when the Saussurean ideas were transmitted to Slavic linguistics, they were absorbed into a fairly balanced discipline in which the synchronic and diachronic approaches complemented, rather than excluded, each other. This tradition continues today. The outstanding Slavists of the twentieth century, scholars like Trubetzkoj, Jakobson and Isačenko, have viewed synchronic processes not only in their own right but also in the context of the nature of historical processes which have formed Slavic. Thanks to them, and to the traditions which they helped to establish, the analysis of the modern Slavic languages is more consistent with the historical perspective. Particularly in phonology and morphology, the historical approach helps to explain much about the structure of the modern languages. It is the historical viewpoint, in short, which provides a key to what Stankiewicz (1986) calls the "unity in diversity" of modern Slavic.

0.7 **Organization**

This book is designed principally for English readers with some competence in descriptive linguistics. It does not assume a knowledge of the Slavic languages, nor of modern theoretical linguistics. We present a typology of the Slavic languages – what makes them a language family, and how they differ from one another – using conventional linguistic notions and terms. Our approach is oriented not towards theoretical problems of linguistics or typology, but rather towards describing and contrasting the modern Slavic languages and their linguistic properties: to show the characteristics of each of the languages in itself, and in its relations to other Slavic languages. The map on p. xx shows the geographical location of the modern Slavic languages.

The selection of data from the Slavic languages presents special problems. Slavic is the most studied language family in the world. In the *Bibliographie linguistique de l'année* for 1998 a raw count of entries shows that out of the 20,743 listings, Indo-European accounts for 53 percent or 11,066 items, and nearly half of Indo-European – 5,370 items, for 25.9 percent of all listings, or 48.5 percent of Indo-European – relates to Slavic. The next most studied language families in order are Germanic (10.8 percent) and Romance (with Italic; 10.1 percent). In the listings for Indo-European, Polish is the most studied language, with 11.4 percent, and on this count is the most studied language in the world. It is followed by Russian with 10.3 percent and – remarkably – English with 8.4 percent, German with 6.3 percent and French with 6 percent. These figures certainly underestimate the position of English in particular, since so much of the material indexed under "General Linguistics" deals directly or indirectly with English. Nonetheless, the domination of Slavic in the world's linguistic research output is evident, although the Slavs account for barely 6 percent of the world's population. Within Slavic linguistics work on Polish accounts for 23.6 percent, and Russian 21.1 percent, a significant fall from the 26–28 percent of a decade ago: apparently the non-Russian members of the former Eastern Bloc are not researching Russian as intensively as they used to. In terms of numbers of native speakers and socio-political position Russian is under-represented, and Polish over-represented. But these are the languages about which most is written, and the selection of material and references in this book often tends to reflect this fact. A typological study, however, has to be sensitive to qualitative as well as quantitative data, and for this reason the presence in the literature of the "larger" languages will be counter-balanced by a special focus on the less represented languages like Sorbian, which present features of inherent typological interest. Many students of the Slavic languages get only a short distance past the large languages, and many no further than Russian. In a survey like this there is also a temptation to spread the examples widely throughout the rich material and literature. We have tried to clump the examples together, so as to avoid buckshot exemplification. Readers will find blocks of examples on Sorbian, say, which serve to emphasize not only individual points of the language's characteristics but also their coherence within the language. There are substantial numbers of Russian examples, since readers are most likely to know Russian if they know a Slavic language; since it is often useful to have a point of reference and comparison in a known language; and since Russian not only outnumbers the speakers of all the other Slavic languages together, but has also had a larger historical, ideological and cultural influence on the other Slavic languages.

0.7.1 Transcription and transliteration

(The orthography and transliteration of all languages is given in appendix B.) Roman-script Slavic languages (Czech, Slovak, Polish, Upper and Lower Sorbian in West Slavic; Slovenian, Bosnian and Croatian in South Slavic) are given in the standard orthography, with the standard linguistic diacritics for tone and length in Bosnian, Croatian, Serbian, and for length and tone or quality in Slovenian, where we need to highlight phonological properties, even though these diacritics belong to dictionaries and specialized use rather than to regular orthography: see 0.7.2.

Languages written in the Cyrillic script (Russian, Ukrainian and Belarusian in East Slavic; Serbian, Bulgarian and Macedonian in South Slavic) are given in Roman transliteration. For Serbian the Roman version of the standard ekavian variant is used throughout, as was standard in the Roman-script version of Serbo-Croatian (2.2.4, 10.2.2, appendix B). For the other languages the transliteration follows the widely accepted "American–European" Slavic linguists' model, which is close to a letter-by-letter transliteration with the exception of the "yotated" vowels (pronounced with a preceding /j/ element) and compound consonant letters like *šč* (see (3b) and appendix B). In (3a) both *ë* and *ja* contain two elements: a vowel nucleus; and a preceding "j" sound which is realized as [j] syllable-initially and after vowels, and as a softening (palatalization) of preceding consonants (here [tʲ]). The Cyrillic "soft sign" (ь), which marks palatalization of preceding consonants, is marked with a following prime accent in the transliteration, and corresponding to the current IPA (International Phonetic Association) convention, the superscript "j" [ʲ] is used in phonetic transcription:

		Cyrillic	Transliteration	Phonetics	Translation
(3a)	Rus	тётя	tëtja	[ˈtʲötʲə]	'aunt'
		яйцо	jajcó	[jɪjˈ1t͡sɔ]	'egg'
(3b)	Rus	женщина	ženščina	[ˈʒɛnʃʲʃʲɪnə]	'woman'
(3c)	Rus	быть	byt′	[bɨtʲ]	'to be'

0.7.2 Accent and stress

The three East Slavic languages (Russian, Ukrainian and Belarusian), as well as Bulgarian, are Cyrillic script and have a word accent which is free (it can occur on any syllable in the word) and mobile (it can move between syllables within paradigms). Macedonian, also written in Cyrillic, normally has antepenultimate stress.

In our orthographic transliteraton word accent is marked with an acute accent on the accented syllable in examples (but not usually in proper names or in the titles of books or articles, unless their phonetic properties are relevant). In phonetic representations – using IPA – we use the conventional vertical bar before the accented syllable:

(4a)	Rus	*rukáˊ*	'hand'	[NomSg]	[rʊˈka]
(4b)	Rus	*rúku*	'hand'	[AccSg]	[ˈrukʊ]

All West Slavic languages are written in the Roman script and have fixed stress. Polish has penultimate stress, and the others have initial stress. None of these languages marks word-stress orthographically. Diacritics are part of the standard orthographies of these languages, as described in appendix B, and indicate modifications of quality in consonants (5a), or quality and/or length in vowels (5b):

(5a)	Cz	*mazat* [ˈmazat] 'to smear'	*mažu* [ˈmaʒu] 'I smear'
(5b)	Cz	*dům* [duːm] 'house'	*domu* [ˈdɔmu] [GenSg]

Occasionally, however, even these languages exhibit unusual stress patterns, like Polish *czterysta* '400', which is stressed, abnormally for Polish, on the antepenultimate syllable instead of the expected penultimate. In such cases we shall mark stress in the IPA style, with a preceding prime:

(6)	Pol	ˈ*czterysta* '400' [ˈtʃtɛrɬsta]

Slovenian, Bosnian, Croatian and Serbian have – at least in some variants of the standard – preserved word-tone and vowel length, the former on the accented syllable only, which can be marked in their orthography, but is printed only in dictionaries and pedagogical texts. For Slovenian the default variant will be the non-tonal one, showing only length and vowel quality (3.5). We follow these conventions:

(7a)	Sln	*séstra* 'sister' [NomSg] (tonal variant: long rising tone on tonic) *sêstra* (non-tonal variant: long low- mid [ɛ])
(7b)	B/C/S	*sèstrē* 'sister' [GenSg] (short rising tone on tonic + post-tonic long vowel, here final)
(7c)	B/C/S	*sȅstro* 'sister' [VocSg] (short falling tone on tonic)
(7d)	B/C/S	*séstrā* 'sister' [GenPl] (long rising tone on tonic + post-tonic long)

0.7.3 Structure of the examples

The examples are selected to illustrate specific points, and to provide comparisons with other relevant data in other Slavic languages. Examples are numbered, with a new numerical sequence in each chapter. If the example deals only with one language, that language is identified immediately after the example number, so that "Sln" stands for "Slovenian" in (8a). If the example covers more than one language, the language of each is identified immediately before each data item, so that (8b) refers to examples from Russian and Slovenian. Translations are provided in single quotation marks. Grammatical information is given in square brackets (8c). Phonetic representations are also in square brackets, and follow the principles of the IPA. Phonetic representations are as narrow as is appropriate for each example (8d). Glosses are given where appropriate, and where the translation does not provide enough guidance to the structure of each word and phrase. Hyphens within a word indicate a morpheme boundary, and are given only where the internal structure of the word is under discussion (8e).

(8a) Sln *séstra* 'sister' (long rising tone)

(8b) 'sister': Rus *sestrá* Sln *séstra*

(8c) 'sister': Rus *sestrá* [NomSg]

(8d) Rus *sestrá* [sɪˈstra] [NomSg] 'sister'

(8e) Rus

sestr-á	*čitá-et*
sister-[NomSg]	reads-[3Sg NonPast]

'(my) sister is reading'

0.7.4 Abbreviations

Language-name abbreviations and abbreviations of linguistic terms are given in appendix A.

0.8 **Outline**

The book begins with the linguistic evolution, genetic affiliation and classification of the Slavic languages – their relation to the Indo-European parent language, and particularly to Baltic, the closest of the language families to modern Slavic (chapter 1). We then turn to the "external history" of Slavic, covering the early history of the Slavic tribes, the emergence of Slavic from the Indo-European language family, and the national, cultural and linguistic formation of the modern

languages (chapter 2). This initial survey takes a more language-by-language approach, in order to highlight the essential characteristics of the three groups of Slavic (South, East and West) and of the languages which they include. The following nine chapters take a more topic-based cross-language perspective, and are concerned with phonology and phonetics (chapter 3), morphophonology (chapter 4), inflexional morphology (chapter 5), morphosyntax (chapter 6), sentence structure (chapter 7), word formation (chapter 8), lexis (chapter 9), dialectology (chapter 10) and selected key sociolinguistic issues (chapter 11). Appendix A covers abbreviations of language names and linguistic terms. Appendix B deals with transliteration and orthography. And appendix C establishes an entry-point into Slavic linguistics for those approaching it from the viewpoint of a non-Slavic training. The bibliography is designed to guide the reader in exploring the wider literature. There is an emphasis on English-language sources, followed by French, German and Russian. Anyone seriously interested in Slavic linguistics will have to learn at least Russian in order to gain access to the very large literature. But there is more than enough material in English to provide the linguist with a sound starting-point, and in some areas (like syntactic and phonological theory applied to Slavic) much of the key literature is written in any case in English, which is becoming, in Slavic linguistics as elsewhere, the default language of international scholarship, especially since the end of Euro-Communism.

1

Linguistic evolution, genetic affiliation and classification

1.1 The Slavs: prehistory

The Slavs, according to archaeological and linguistic evidence, can be traced back to around 4000 BC. At this time the Great Eurasian Plain was inhabited by the people (or peoples) whom we now know by the name "Indo-Europeans". Our factual knowledge of this distant period of European pre-history is sketchy and partly conjectural. Although we can only guess how far their territory extended, it is possible that at least the European center of the Indo-European homeland – if not the original homeland itself (on one widely held view) – was in what is now Western Ukraine, and that they spoke a fairly homogeneous language.

By about 3000 BC the Indo-Europeans had occupied most of Europe. What had previously been local dialect variations in the Indo-European language would have begun to diverge in the direction of separate languages (based on linguistic difference), which led ultimately to the now familiar Indo-European language families.

The emergence of Proto-Slavic occurred around 2000–1500 BC. This is the period of Proto-Slavic unity, when the Slavs inhabited a broadly coherent land area, though its exact location remains a matter of some controversy. According to Birnbaum's (1979: ch. 1) summary of the then current state of play, the Slavs' first homeland was north of the Carpathian Mountains, and possibly to the east of the Carpathians' westernmost extremity in the Sudeten Mountains, both of which approximately mark the border between present-day Poland and the Czech and Slovak Republics. An alternative, and perhaps more popular, view favors a location further east, on the middle Dnieper (Schenker 1995: 7 – who offers a good discussion of the various theories). By the fourth century AD the Slav area stretched from the Oder (Pol *Odra*) River in the west to the Dnieper (Rus *Dnepr*, Ukr *Dnipró*) in the east. In the north they had reached the Masurian Lakes in central Poland, the Baltic Sea and the Pripet (Pol *Prypeć*; also Eng *Pripyat*, from Ukr *Prýp'jat'*) Marshes. During this period the Slavs would have spoken a fairly uniform language. Although dialect differences soon began to appear, resulting

inter alia in the division into Baltic, Slavic or an intermediate Balto-Slavic, the pace of linguistic change was relatively slow.

The Slavs' original language is called both "Common Slavic" (Rus *obščeslavjánskij jazýk*, Ger *Gemeinslavisch*, Fr *slave commun*) and "Proto-Slavic" (Rus *praslavjánskij jazýk*, Ger *Urslavisch*, Fr *protoslave*). The difference is partly a matter of national convention and preference, and partly a matter of definition: some scholars assign the name "Proto-Slavic" to the earlier period, and "Common Slavic" to the period preceding the breakup of Slavic unity in the first millennium AD. As Birnbaum notes (1979: ch. 1), it is not possible to determine a clear separation between these periods, so that one might as well use a single term to cover the whole period, with the additional labels "early" and "late" where needed. While Birnbaum opts for "Common Slavic", we shall follow Comrie and Corbett (1993) in opting for "Proto-Slavic" (abbreviated as 'PSl').

According to general consensus in what is still a controversial area, the real break-up of Proto-Slavic unity began about the fifth century AD. There seems to have been a steady expansion to the north and east by the Eastern Slavs. For the others there is evidence that their migrations were related partly to the disintegration of the Roman and Hun empires and the ensuing vacuum in Central Europe. One group of Slavs moved westwards, reaching what is now western Poland and the Czech Republic, and the eastern and north-eastern part of modern Germany. A second wave broke away to the south towards the Balkan Peninsula, where they became the dominant ethnic group in the seventh century, some (in the east) in turn being conquered by the Bulgars, a non-Slavic people of Turkic Avar origin. Of the original Bulgars we have only a few loan-words as monuments, since they were linguistically and culturally assimilated to the Slavs. This group of Slavs – the future Bulgarians and Macedonians – would most likely have moved south via the Black Sea coast, to the east of the Carpathians, meaning that they were in origin Eastern Slavs; this may account for certain typological similarities between Bulgarian and East Slavic (see 1.4).

The Slavs, however, were not politically unified or well organized, and elected a joint leader (or "great prince", using a word [Rus *knjaz′*] derived from the Germanic form of 'king') only when external danger forced them to. This lack of a centralized power structure was to cause serious problems for all the Slavs in their new homelands. It also contributed to the regionalization and differentiation of the emerging Slavic dialects, which by about the tenth century AD had divided into the three major sub-groups of modern Slavic: the South Slavs, East Slavs and West Slavs.

We shall concentrate on three points in the evolution of Slavic: the status of Slavic within the Indo-European group (1.2); the relative standing of Slavic and Baltic in the context of the groups which derived from Proto-Indo-European (abbreviated 'PIE': 1.2.1); and the conventional three sub-groups of the Slavic

languages mentioned above (1.3–1.4), which then formed the basis of the further diversification in the evolution of the modern languages (1.5–1.7).

This chapter will set out the genetic linguistic map of Slavic. It will also be our last major excursion outside Slavic in this book. The following chapters (2–11) will concentrate on the unity of Slavic, and the diversity which we find in its individual languages, both with respect to Slavic as a whole, and with respect to the major sub-groupings of languages within it.

1.2 Slavic in Indo-European

The Slavic languages belong to the "satem" group of Indo-European languages, which also includes the Baltic, Indic and Iranian languages. This well-known classification derives from the way each language treats the PIE initial palatal **k'*- of the word **k'm̥tom* '100'.

We shall follow the convention of using asterisks for unattested forms within the text, but not in tables, which they tend to clutter. All Proto-Indo-European and early Proto-Slavic forms are unattested. Where late Proto-Slavic forms are attested in Old Church Slavonic (1.5.1, 2.2.1) they are not marked with an asterisk.

The "satem" languages (named after the Avestan word for '100') all have a fricative or affricate sibilant for the original **k'* (1a). In contrast, the "centum" languages, which take their name from the Latin for '100', preserve an initial velar stop or its later reflexes (1b):

(1a)	'hundred'	
	PIE	*k'm̥tom*
	Avestan (Old Persian)	*satəm*
	PSl	*sŭto* (*sъto*)
(1b)	Lat	*centum*
	Gk	(*he*)*katón*
	Germanic/Gothic	*hund*(*ert*)

1.2.1 Slavic and Baltic

One of the more contentious issues in the early history of Slavic concerns the nature of the relation between Slavic and Baltic. Clearly, both had a common ancestor in Proto-Indo-European. Equally clearly, the Baltic languages (Latvian, Lithuanian and the now extinct Old Prussian) form a later group which is distinct from Slavic. But did the Baltic and Slavic languages break away from Proto-Indo-European as a single conjoint group, which later divided into Slavic and Baltic?

Or did they break away from it in two separate groups at roughly the same time, and then develop further similarities through geographical contiguity, parallel development and interaction, rather than by being part of a single joint proto-language?

The idea that Baltic and Slavic did indeed share a common ancestor language after the breakaway from Indo-European is called the "Balto-Slavic Hypothesis". One of its originators was Brugmann, in his *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen* (Vol. 1) of 1886. And among its best known critics was Meillet, whose *Les dialectes indo-européens* was published in 1908. There is certainly strong evidence to link Slavic more closely to Baltic than to any other Indo-European language family. Whether, however, this is enough to justify the analysis of a Balto-Slavic period of unity is still controversial.

Many of the commonly accepted features which mark the Slavic-Baltic parallelisms are shared, at least in part, with other IE languages. They include the following (with indicative examples).

1.2.1.1 *Phonology*

For the meaning of all symbols and abbreviations, see appendix A.

1. PIE palatal stops become sibilants (part of the *satem* development, see above), probably via affricates:

	PIE	OCS (Late PSl)	Lith
k' > (*t*)*š* > PSl *s*, Balt *š*	*k'm̥tom* 'hundred'	*sъto*	*šimtas*

2. Deaspiration of aspirated voiced stops (also Germanic):

$d^h > d$	*dhūm-* 'smoke'	*dymъ*	*dumai*
$g^h > g$	*g^hord-* 'town'	*gradъ*	*gardas*

3. Delabialization of labialized velars:

$k^w > k$	*k^wŏ* 'what'	*kъto*	*kas*
g^{wh} (via g^w) > *g*	*snoi̯g^wh-* 'snow'	*snĕgъ*	*sniegas*

4. PIE vocalic *r̥ l̥ n̥ m̥* yield *ir/ur, il/ul, in/un, im/um*

l̥ > il ul	*(u̯)l̥k-* 'wolf'	*vьlkъ*	*vilkas*
m̥ > im um	*dek'm̥* 'ten' (*-ĭmt->*)	*desętь*	*dešimt*

5. *tt/dt > st* (see 1.3.1.1–1.3.1.2, under simplification of clusters and opening of syllables):

tt > st	*met-ti* 'throw'	*mesti*	*mesti*
dt > st	*ved-ti* 'lead'	*vesti*	*vesti*

6. Changes to /s/: different results, appearance of /x/ in PSl (also Iranian and Indic languages, e.g. Sanskrit, also with different results):
   s > *x* (PSl) or *š* (Balt) between *k-, r-* and V:

aksi 'axle'	(*oks-* >) *osъ*	(*akš-* >) *ašis*
u̯r̥s- 'top'	*vъrxъ*	*viršus*

   PSl only, not Balt:
   s > *x* between *i-, u-* and V:

ōu̯s- 'ear'	*uxo*	*ausis*
ōi̯su̯ LocPl ending	*-ěxъ*	*-uose*/*-yse*

7. Fusion of /a/ and /o/ (also Germanic):
   Short /ă/ and /ŏ/ merged into /ₒă/ (or /ă/) in both, long /ō/ and /ā/ merged into /ₒā/ (or /ā/) in PSl only (later PSl /ₒă/ > /o/, /ₒā// > /a/):

/ă/	*ăk'si-* 'axle'	*osъ*	*ašis*
/ŏ/	*ŏk-* 'eye'	*oko*	*akis*
/ā/	*b*^{h}*rātr-* 'brother'	*bratъ*	*broter-elis*
/ō/	*dō-* 'give'	*dati*	*duoti*

In levels other than phonology, rather than produce examples involving Baltic, we refer simply to the relevant chapters containing the Proto-Slavic forms.

1.2.1.2 *Morphology and morphophonology*

1. Use of the original PIE ablative for the genitive singular in *o*-stem nouns (5.5.1.1)
2. Declension of active participles (present and past), and extension of their stems by /j/, giving *-Vnt-* > *-ǫ /ę-t'-* for the present and *-u̯s-* > *-vš-* for the past (5.5.5.10)
3. Parallelisms in the formation of some 1 Person singular pronouns without gender (oblique cases) (5.5.3)
4. Shifting the *consonant*-stems into the *ĭ*-stems (5.5.1.6)
5. Separation of the present-tense and infinitive stems as the basis for conjugation (5.5.5.2)

1.2.1.3 *Word formation*

1. Formation of collective numerals with the suffix *-er*/*-or* (5.5.4.3, 8.6.1)
2. Formation of "definite" long-form adjectives by adding a pronoun to an adjective stem (5.5.2.1)
3. Parallel adjectival suffixes, e.g. *-ĭn-, -ĭsk-* (though the latter is general PIE, cf. Eng. *-sh*) (8.4.3)

1.2.1.4 *Syntax*

1. The use of the genitive for the accusative in negated direct objects (6.2.3–6.2.4).
2. The use of the instrumental in the predicate of copulative sentences (*ibid.*, 7.2.2.2)

1.2.1.5 *Lexis*

Uniquely shared lexical items (Trautmann, 1923):

(2)

	PSl	OCS	Lit	B/C/S	Rus	Pol
'head':	golva	glava	galvà	gláva	golová	głowa
'linden':	lipa	lipa	líepa	lȋpa	lípa	lipa
'horn':	rogŭ	rogъ	rãgas	rôg	rog	róg
'hand, arm':	ronka	rǫka	rankà	rúka	ruká	rękа

Many Slavists (e.g. Horálek, 1955: 71; Shevelov, 1964: 613) have taken the line that one does not need to take a strong position on the Balto-Slavic Hypothesis in order to do viable Slavic linguistics, and we shall follow their example.

1.2.2 Slavic and other Indo-European language families

1.2.2.1 *Slavic and Iranian*

The Slavs were in direct contact with the Iranians in the south from about the seventh century BC to the second century AD. Not enough is known of the phonology of Scythian or Sarmatian, the Iranian languages in closest contact with the Slavs, but some broad-scale phonological correspondences between Slavic and Iranian (and often with Baltic as well) can be established.

A phonological feature which may stem from this contact is the change of /s/ to /x/ in certain contexts (between /i, u, r, k/ and V; also known as the *ruki* rule). We have seen above (1.2.1) that Baltic (and also Iranian and Indic) apparently shared part of this change (Baltic after /u, k/ only), and also that the Baltic result was /š/, not /x/. It could well be that the frequency of /x/ in Iranian languages contributed at least to the last step in Proto-Slavic, as it provided other sources of /x/ also, e.g. Rus *xoroš-* 'nice', possibly from the Iranian root *xors-* 'Sun God' (but refuted by Vasmer, 1964–1973, s.v. *xorošij*).

1.2.2.2 *Slavic and Germanic*

The Slavs were in contact with the Goths from the second to fourth centuries AD during the 'wanderings' of the latter as far as the Black Sea. Later the West Slavs

were to have close contacts with the Germans, but that was to be at the time of the Proto-Slavic break-up, so that the effects are local, and reflected only in some Slavic areas, notably Sorbian, but also Czech, Slovak and Slovenian.

1.3 **Proto-Slavic**

"Proto-Slavic" (see above) is the name given to Slavic after its separation from Indo-European and Baltic, but before the partition of Slavic into its three main sub-families. This period lasted probably for more than 2,000 years, and the features listed below belong to what is assumed to be the core of the common language.

The earliest phonological developments, centered on suprasegmental features, were:

1. tone ceased to be phonemic, becoming a concomitant of length: long vowels and diphthongs had rising pitch (known as "acute"), and short vowels had falling (non-rising) pitch (known as "circumflex");
2. then the nuclear vowels of diphthongs were shortened, but the tone distinctions remained, so that tone again became phonemic, but only on diphthongs;
3. still later, word-stress arose from tonal features: an acute (or the first acute in a word) became the stressed vowel. If there was no acute, the word was stressless, and initial stress developed on such words in most areas. The pitch on endings thus accounts for the mobile patterns which developed. Inflexions or suffixes with acute following circumflex stems account for fixed-end patterns.

The probable shape of the early Proto-Slavic phonological system is shown in table 1.1. Bracketed forms joined by a vertical line are likely regional variants; the vowel symbols ₑă etc. represent notional intermediate phonemes. These reflect partly the dual origin of these phonemes, especially the back set, but mainly point to their future development at the time when length was replaced by quality (1.3.1.6). The use of the term 'diphthong' to cover sequences which are structurally 'vowel + sonorant' in pre-consonantal position, a much broader application than is normal elsewhere, is common within Slavic linguistic usage. In particular, it facilitates the description of the development of such sequences.

Developments within Proto-Slavic The following discussion is topic-based rather than chronological, though as far as possible chronological order is followed. For clarity of exposition the vowel symbols used from here on are those of later Proto-Slavic. The equivalences are shown in table 1.1 and discussed in 1.3.1.6.

Table 1.1. *Early Proto-Slavic phonological system*

Consonants

	Labial	Dental	Palatal	Velar
Stop	p	t		k
	b	d		(g)
Nasal	m	n		
Fricative		s		x
	(v)	z	j	(ɣ)
	(w)			
Liquid		r		
		l		

Vowels (in brackets are given the later Proto-Slavic symbols used from 1.3.1 on)

	Front	Central	Back
High	ĭ ī (ъ i)		ŭ ū (ъ u)
Low	ₑă ₑā (e ě)		ₒă ₒā (o a)

Diphthongs (pre-consonantal VOWEL + SONORANT)

Low short vowels plus i̯, u̯: ₑăi̯ ₒăi̯ ₑău̯ ₒău̯(ei̯, oi̯, eu̯, ou̯)
All short vowels plus, r, l, m, n, e.g. ĭr ₑăr (ьr, er)

Suprasegmental

Stress	Free
Quantity	On all monophthongs (not on the above diphthongs)
Tone	Automatic acute on long vowels; diphthongs may have acute or circumflex

1.3.1 Phonology

1.3.1.1 *Syllable*

In the following discussion of syllable structure we interpret the 'boundary' empirically, that is, as reconstructed on the basis of known developments and not on any abstract theoretical notion.

A generalization which may be made about almost all developments – and certainly all the major ones – within Proto-Slavic is that the syllable boundaries and relations within the syllable altered: the unit of the syllable became more discrete, the boundary was marked by a drop in sonority. This is the so-called 'Law of Rising Sonority' (within the syllable), which led to the 'Open Syllable Law', that is, syllables became 'open', always ending in a vowel (see below on word-final consonant loss). Initial consonant clusters were acceptable to the extent that they conformed to this rule (e.g. fricative + stop + glide, as *str-*, *zgl-*). Further, all the elements within the syllable would have been closely bound and able to influence one another, for instance to cause assimilation.

The first implementation of this syllable restructure, the 'opening of syllables', had major consequences for all aspects of the phonological system. We shall see the main effects below, but probably one of the earliest was the dropping of word-final consonants (e.g. PIE NomSg endings *-ŭs/ŭm* etc. > PSl *-ŭ/-ŏ*). Also early was the simplification or reorganization of unacceptable consonant clusters (usually of falling sonority, see 1.3.1.2).

1.3.1.2 *Simplification of clusters*

The opening of syllables in Proto-Slavic by elision (i.e. deletion) of syllable-final consonants was an effect of the more general principle of 'rising sonority': in accordance with this principle inherited consonant clusters remained acceptable only if they conformed to it. Otherwise some sort of simplification took place: elision, dissimilation or metathesis. Thus geminate clusters were always simplified. This is probably the earliest such simplification, as it is shared by Baltic (see above), and similar effects are observed in Greek and Iranian, either by elision (e.g. fricatives, as in NomSg final *-s*) or by dissimilation (e.g. stops: *tt* > *st*: PSl **plet-ti* 'braid, plait' Inf > *plesti*, *dt* > *st*: *ved-ti* 'lead' Inf > *vesti*). Clusters of 'stop + any consonant' usually involved the elision of the stop, e.g. PIE **wapsa* 'wasp' > PSl **osa* (Lith *vepsva*); **greb-ti* 'row, rake' Inf > *greti*; **greb-s-* (the aorist of the same verb) > *grěs-* (with lengthened root vowel). However, clusters of fricative + stop (e.g. *st*, *zg*) were accepted (and were indeed used as solutions to 'stop + stop', see *plesti* above) even though the stop was less sonorous, which indicates that ease of articulation within the overall open syllable structure was sometimes in conflict with rising sonority.

When a non-sonorant was followed by a sonorant, the first of the two consonants was usually kept, since such a cluster did comply with rising sonority (e.g. *tri* 'three'). Clusters of 'fricative + sonorant' were always accepted, on the same grounds as 'fricative + stop', cf. word-initial *sn-* (*sněg* 'snow'), *zn-* (*zna* 'know'), *sl-* (*slov* 'word'), *zv-* (*zvon* 'ring'). While 'stop + sonorant' clusters were also normally accepted (*tri*), the clusters *tl* and *dl* show hesitation by area, producing one of the distinguishing features of West Slavic, which alone retained them (1.7.1).

1.3.1.3 *Syllabic harmony*

Within the new syllabic structure (C(C)V) the relationship between the consonant and the following vowel would have become very close, such that each influenced the other. This general feature is known as 'syllabic harmony' (sometimes 'synharmony', Rus *singarmónija*). The features affected are centered on front/back

tongue position, which for vowels means indeed front/back, and for consonants means the raising of the tongue at the front or back. Thus we have effects like the following: (1) a front vowel caused raising of the tongue at the front during the consonant articulation, producing a 'palatalizing' effect, and so-called 'palatalized' or 'soft(ened)' (or 'sharp') consonants; (2) in the case of the velars, the result was their conversion to full palatal consonants, that is, ones articulated in the high front tongue position (a shift from soft to hard palate location); (3) where palatal consonants had arisen as a result of the effect of [j] (see below), and where the following vowel had been a back one (say, /o/), this back vowel was fronted by the influence of the consonant. In the case of short /o/ and /ъ/ the fronted versions were, in fact, the front partners /e/ and /ь/ (e.g. PSl *měst-o* 'place' ~ *pol-e* 'field', *rab-ъ* 'slave' ~ *kon-ь* 'horse'), suggesting that any rounding was not strong in these short vowels. On the other hand, for the long /u/ (< ū) and nasal /ǫ/ (< 'back vowel + nasal consonant' before a consonant, see 1.3.1.7) the result was at most a fronted [ü] or nasal [ǒ], that is, they did not lose their rounding and merge with their front 'partners' (/i/, /ę/). Most interesting is the long /a/ (< $_{o}$ā), whose formal front partner was /ě/ (< $_{e}$ā) (e.g. PSl *čaša* 'cup'): it seems that a fronted [ä] was the most common result, and since this is what we believe the pronunciation of /ě/ itself was in most areas, this result reflects normal fronting. However, the ultimate result in all areas was a reflex of /a/, and not of /ě/ (e.g. not **češe*), and moreover, the reflex of /ě/ itself preceded by a palatal was also /a/ (!) (e.g. PSl **ležěti* 'lie' > *ležati*). Further, in those areas where /ě/ in other contexts shifted to another position (higher, e.g. in East Slavic) it was not joined by the vowels after palatals (e.g. PSl *lěto* > Rus *léto*, but Rus *ležát'*), that is, the [ä] of these areas (*ležäti*) was not identical to /ě/. Incidentally, no area has retained the three fronted articulations (*ä, ü, ǫ̈*), which were only allophones, and which ceased to function after the syllable was again restructured – at which stage the sequence 'palatal consonant + back vowel' again became acceptable (see 3.2.1).

The extent to which these shifts are reflected in the modern languages varies: for some (e.g. Russian, Polish) the "soft" articulation of consonants – and not just before front vowels – has become an inherent feature; and for all languages the fronting of vowels is reflected in morphophonological alternations, in particular in the opposition between hard and soft declension types (e.g. Russian neuter nouns, hard *mést-o* 'place' *vs* soft *pól-e* 'field', see 4.3).

1.3.1.4 *Palatalization by [j]*

An early consonant change which was to be a major feature of Slavic was the production of palatal consonants by 'fusion' with a following [j] (in reality palatalization by [j] followed by loss of the [j]). Sequences of 'consonant + *j*' would

always have been acceptable within the new syllable unit, so this change is not tied to the syllable structure. But it does seem probable that it is a manifestation of an early (regressive) assimilative tendency which could be the precursor of the opening of syllables.

There are discrete stages in this development. Most subsequent results are common to the whole group, but some are not. In order better to observe the process, it is useful to see the first stage as fused consonants which are simply palatalized, or jotated, versions of the base consonant, e.g. stops like *p'*, *t'*, *k'*, fricatives like *s', x'* and sonorants like *r', l', n'*.

In a few cases consonant clusters were affected as a unit, for example the clusters *kt* and *gt* jointly became *t'* at this first stage.

The final stage, producing consonants like the palatals or post-alveolars, is late, occurring after the syllable opening. Some of the results are phonetically "natural", and occur in many languages, if only in allegro speech, e.g. *s' > š'*, *z' > ž'* (cf. Eng *sure, pleasure*), common to all the group. Others are less so, for example the labials become clusters of labial plus *l'* – that is, the palatal element is realized as a palatal lateral [ʎ]. There is no agreement about whether this *l'* element was initially common or not. The western group (e.g. Polish, Czech) do not show it except in a few odd cases, so it is possible either that they never developed it and borrowed those few cases, or that they subsequently lost it and these cases are remnants (see 3.2.2). For the sonorants *r, l, n* there is no special further development until after the break-up into East, West and South Slavic. The velars show mostly common results: all languages have *k' > č'* and *x' > š'*:

(3a) 'I cry': PSl **plak-j-om* > *plačǫ*: all except Sln have the stem *plač-* (Sln *plakam* by analogy with Inf)

(3b) 'soul', PSl **dux-j-a* > *duša*: Cz *duše*, all others (including OCz) *duša*

The results of *g'* may have depended on the stop or fricative nature of this sound, the stop [gʲ] giving *dž'* (locally later simplified to *ž'*), and the fricative [ɣʲ] giving *ž'*. If the fricativization of /g/ is placed later – which is the traditional view – it would be accompanied at that later time by the deaffrication of *dž'* to *ž'*. Positing the early local occurrence of fricative /ɣ/ thus would seem reasonable.

1.3.1.5 *Velars and their palatalization*

A change which seems plausibly to predate the syllable opening is the fronting of the velars in the vicinity (on either side) of a front vowel, e.g. BSl **k'etūr-* 'four' > PSl **k'etyr-*; **-ĭkŭ* 'agent suffix' > **-ьk'ъ*. Within the revised (open) structure mutual effects across a sonority boundary should be minimal (and those

within the unit maximal), so that the influence of a vowel on the following consonant is less likely. The second example (*-ьk'ъ) might well therefore have been in place before the restructuring. We shall see below that this particular development of the velars has caused problems of chronology, and the positing of an early stage of fronting across the syllable boundary while it was still weak can help with this problem.

The above results of the effect of [j] on consonants may now be expanded to the specific context of the new open-syllable structure. Vowel articulation may now influence the articulation of preceding consonants: in particular, we would expect the typical palatalizing effect of a front vowel, producing fronted allophones. This is effectively the result for dentals and labials. However, for velars the result was much more drastic, matching the effect of /j/ noted above: again, in Proto-Slavic all areas have *k' > č'*, *x' > š'* and *g'/ɣ' > ž'* (possibly via *dž'*). These changes are traditionally called the '1st Palatalization of the Velars' (PV1), and the context is: before any of the front vowels existing at that time, namely high /i/ and low /e/, which may both be long or short (these are the vowels which became /i ь ě e/).

(4a) 1st Palatalization of the Velars (PV1)

PSl	B/C/S	Rus	Cz
(*kětūre >) *k'etyre > četyre* 'four':	*čètiri*	*četýre*	*čtyři*
(*gīv- >) *g'iv- > živ-* '(a)live':	*žîv*	*živój*	*živý*
**strax'-ьn- > strašьn- 'terrible':*	*stráš(a)n*	*strášnyj*	*strašný*

This change occurred in important inflexions, like the present tense of verbs, with the thematic vowels *e* and *i*:

(4b)

PSl	B/C/S	Rus	Cz
**pek'-e-tь > pečetь* 'he bakes':	*pèčē*	*pečët*	*peče*
**mog'-e-tь > možetь* 'he can':	*mȍžē*	*mózet*	*může*
**dyx'-e-tь > dyšetь* 'he breathes':	*dȉšē*	*dýšit*	(arch.) *dýše*

After this palatalization of the velars, producing post-alveolar consonants (/č/, /š/, /ž/), there occurred a second fronting process affecting the velars, producing probably first (alveolo-)palatal sounds like those in modern Polish and B/C/S, which by the time of the break-up had become soft dentals or alveolars. This is the reverse of the actual Polish and B/C/S cases, where these palatals were derived from soft dentals:

/k/ > /c′/, /g/ > /dz′/ or /z′/, /x/ > /s′/ or /š′/ (examples in (5)).

The variants for /g/ may have been parallel to those for the first set, that is, they may have related initially to the stop *vs* fricative nature of /g/. However, subsequently, in the same way, even areas with stop /g/ converted /dz′/ to /z′/. The variants for /x/, on the other hand, are geographically based, the post-alveolar /š/ occurring in the West only.

The context for this set of velar frontings is twofold. One is the expected following front vowel, specifically the two new front vowels /ě$_2$/ and /i$_2$/. They had not been present at the time of the first change, but had arisen later from the (monophthongized) diphthong /oĭ/ (see 1.3.1.7), and had merged with existing /ě/ and /i/. This set of changes is referred to as the '2nd Palatalization of the Velars' (PV2):

(5a) 2nd Palatalization of the Velars (PV2)

PSl	B/C/S	Rus	Cz
(**koi̯na* >) **k'ěna* > *c'ena* 'price':	*céna*	*cená*	*cena*
**goi̯lo* > **g'ělo* > *zělo*: 'very'	Sln *zeló*	RChSl *zeló*	OCz *zielo*
**xoi̯d-* > **x'ěd-* > *sěd*/*šěd-* 'grey (haired)'	*sêd*	*sedój*	*šedý*

This change, too, occurred in important inflexions, e.g. the locative singular and nominative plural of nouns, and the imperative of verbs (though most often this alternation has not survived analogical levelling):

(5b)

PSl	B/C/S	ORus	Cz
(**ronk-oi̯* >) **rǫkě* > *rǫcě* 'hand' LocSg	*rúci*	*rucě*	*ruce*
(**nog-oi̯* >) **nogě* > *nodzě* 'foot' LocSg	(*nòge*)	*nozě*	*noze*
(**doux-oi̯* >) **duxě* > *dušě*/*dušě* 'spirit' LocSg	*dùsi*	*dušě*	*duše*

The second context in which this same set of changes occurred is more complex. It appears to be caused by a *preceding* high front simple or nuclear vowel (that is, mainly long and short /i/) (examples in (6)). This makes it a progressive assimilation, which in itself suggests rather the period *before* the opening of the syllables, since after that there was a clearer boundary between a vowel and a following consonant. On the other hand, the fact that only one diphthong with nuclear [i] – '*i* + nasal sonorant' (but not '*i* + liquid sonorant') – provokes the change means that it almost certainly occurred after the quantity changes (1.3.1.6) and monophthongization (1.3.1.7). (The *i* + *N* diphthong would have become first a nasalized [ĩ], then merged with the lower nasalized /ę/ [ɛ̃].) The identical results to the 2nd Palatalization also suggest a similar late period.) These conflicting facts have led to a range of interpretations about this set of velar changes. Traditionally it has been called the '3rd Palatalization of the Velars' (PV3), suggesting a late chronology,

but it is now more commonly called the 'Progressive Palatalization of the Velars', and many scholars place it as the earliest of the three. A compromise position, which attempts to accommodate the contradictions, sees it as having occurred in two stages: the first is early, and produced simply fronted velars; and the second is simultaneous with PV2, taking these sounds along with those in the new front vowel context (PV2) forward to the palatal area. This is why we favor the view suggested above of the early appearance of fronted velars when adjacent to front vowels.

There are also several other complications with this set of changes. One is that certain following vowels (high and/or rounded) seem to have prevented it, which fits with the later (open) syllable situation. Second, there must have been some analogical levelling, for example in a paradigm where the following vowel was sometimes a preventer, sometimes a supporter (mainly /a/), and analogy would be particularly strong in this type, where the motive force was never an inflection, but always a stem vowel. These last facts account for the absence of alternations arising from this set:

(6) 3rd (Progressive) Palatalization of the Velars (PV3)

PSl	B/C/S	Rus	Cz	Pol
(*ot-ĭkŭ >) *ot-ьk'ъ > ot-ьc'ь 'father'	*òt(a)c*	*ot(é)c*	*ot(e)c*	*ojciec*
(*kŭnĭngŭ >) *kъnęg'ъ > kъnędz'ь 'prince'	*knêz*	*knjaz′*	OCz *kněz*	*ksiądz*
(*vĭxŭ) > *vьx'ъ > vьs'ь 'all'	Sln *v(è)s*	*v(e)s′*	OCz *v(e)š*	OP *wszy*
suffix *-īkā > *-ik'a > -ic'a	*-ica*	*-ica*	*-ica*	*-ica*

Note that the former following back vowels which did not prevent the change were then fronted by the rules of syllabic harmony (1.3.1.3), as in the final vowel of the first three examples of (6).

1.3.1.6 *Vowel quantity > quality*

The system of four pure vowels with long and short versions was replaced by a system in which quantity ceased to be distinctive and was replaced by qualitative distinctions. Length was probably preserved phonetically for some time and was automatic, or residual, in the new vowels. We state the starting and finishing-points ('Early' and 'Late') in table 1.2:

On the quality of the new vowels, we can see that:

a. rounding has become distinctive in the low vowels. But for the high vowels, on the contrary, rounding was weak and was to be lost from

Table 1.2. *Vowel quantity > quality in PSl*

Early PSl		Late PSl		B/C/S	Rus	Cz
$_{o}$ă	*n_{o}ăktĭ	o	*not'ъ* 'night'	*nôć*	*noč′*	*noc*
$_{o}$ā	*d_{o}ā-tĭ	a	*da-ti* 'give'(Inf.)	*dȃti*	*dat′*	*dát*
$_{e}$ă	*m_{e}ădŭ	e	*medъ* 'honey'	*mẽd*	*mëd*	*med*
$_{e}$ā	*s_{e}ād-	ě ([ä])	*sěd-*'sit'	*sed-*	*sed-*	*sed-*
ĭ	*vьĭx'ŭ	ь ([ĭ])	*vььs'ь* 'village'	Sln *vâs*	*ves′*	*ves*
ī	*gx'īvŭ	i	*živъ* 'alive'	*živ*	*živój*	*živý*
ŭ	*dŭkt(er)ī	ъ ([ŭ] or [ə])	*dъt'(er)i* 'daughter'	*kćî*/*kćêr*	*dočʹ(-er)-*	*dcera*
ū	*dūmŭ	y ([ɨ])	*dymъ* 'smoke'	*d'ȉm*	*dym*	*dým*

the long /u/, and by inference also from the short. The subsequent development of /ъ/ leaves its rounding status at this stage unclear. Since strong rounding of the high vowels appeared only with the monophthongization of diphthongs, i.e. when /ou̯/ > /u/ (1.3.1.7), it is also possible that it was only then that rounding was removed from the original long and short /u/.

b. the short high vowels are assumed to have lowered to high-mid and more centered. In other words their quality may be said to have been 'reduced' (from periphery to center), so that they are frequently referred to as the 'reduced vowels' (Rus *reducírovannye*). Their symbols are the Cyrillic letters used for these vowels in OCS – ъ (back) and ь (front); see also (d) below on their length.

c. the quality of /ĕ/ is also a matter of disputation, again because of the variety of later reflexes. Most popular is the view that at this stage it was a low front unrounded vowel ([ä] or [æ]). For others it was a rising diphthong of the [i̯a]/[i̯ä] type. And for still others, it had already shifted to a higher position ([e] or [i]) in some areas, namely East Slavic. This last position is intended to account mainly for local reflexes, see 3.2.1.3, but the evidence from borrowings from East Slavic, e.g. into Finnish, suggests a change of quality in the East: *měra* 'measure' is borrowed early as Finn *määrä*, while *věstь* 'news' is borrowed later as Finn *viesti*. Its symbol is the Czech letter *ě*, representing its common reflex in Czech.

d. the quantity of the old vowels continued to reside in the new vowels, so that /a/, /ĕ/, /i/ and /y/ were residually long, and the other four short. The subsequent developments of /ъ/ and /ь/ suggest that they were even shorter than /o/ and /e/. But this remains a hypothesis,

related presumably to their higher position; the further shortening may have occurred later. The loss of rounding of /ъ/ causes it to tend towards *schwa* ([ə]), which it becomes in many areas (3.2.1.1).

1.3.1.7 *Monophthongization*

Diphthongs (as defined above, 1.3) were prime casualties of the principle of rising sonority, since in principle the semivowel or sonorant second part of a falling diphthong represents a drop in sonority and hence means a closed syllable – at least when followed by a consonant or word boundary. Virtually all such tauto-syllabic diphthongs were "monophthongized" in Proto-Slavic. We say 'virtually', because it appears that the final stages of this process were overtaken by other changes which reversed the syllabic structure and reinstated closed syllables within the system. The diphthongs which failed to complete the process were some of those in which the closing element was one of the sonorants *r, l*, while the rest changed consistently. The results are shown in (7a–c). Note that the vowel nucleus of the diphthong is always short, but the new pure vowel is (phonetically) long:

Diphthongs ending in a semivowel (*i̯, u̯*)

(7a) monophthongization of diphthongs ending in a semivowel (*i̯, u̯*):

PSl	OCS	B/C/S	Rus	Cz
oĭ > ě, i	*berěte, beri* 'take' Imper. 2p, 2s	*bèri*(*te*)	*berí*(*te*)	*ber*(*te*)
eĭ > i	*iti, idǫ* 'go' Inf., 1SgPres	*ìći, ìdēm*	*idtí, idú*	*jít, jdu*
ou̯ > u	*uxo* 'ear' NomSg	*ȕho/ûvo*	*úxo*	*ucho*
eu̯ > (j)u	*l'ud-ъ, -ьje* 'people' NomSg, Pl	Sln *ljȗd-i*	*ljúd-i*	*lid-é*

In the vowel system, the new /ě/ and /i/ merge with the existing ones from long front vowels (table 1.2) (though often they are marked for etymological purposes as /ě$_2$/ and /i$_2$/). The new /u/ occupies the place of the /u/ lost by unrounding to /y/ (and may have given the final push to its unrounding). The front /e/ of *eu̯* is reduced to [j], with its usual effect of fusing with the preceding consonant to make a new palatal consonant.

Diphthongs ending in a nasal sonorant (*m, n*)

(7b) monophthongization of diphthongs ending in a nasal sonorant (*m, n*):
(F = Front, B = Back, N = Nasal, V = Vowel)

		PSl	OCS		(Other PIE)
FV (e, ь) + m/n > FNV	ę	**pentĭ*	*pętь*	'five'	(Gk *pente*)
		**dĕsemtĭ*	*desętь*	'ten'	(Lat *decem*)

BV (o, ъ) + m/n > BNVǫ	*pontĭ	pǫtь 'way'	(Lat *pont-*)
	*domti	dǫti 'blow' Inf.	(Lith *dumti*)

As expected, the reflex of these is a nasal vowel, with the front or back quality of the nuclear vowel retained. The traditional marker of nasality in Slavic linguistic usage is the subscript hook, taken from the Polish alphabet. The underlying symbols are those of the basic (new) mid vowels *e* and *o* (while modern Polish uses underlying *a* for the back nasal – *ǫ*). Since the further development of the nasal vowels varies by area, modern examples are not given here, but in chapter 3 (3.2.1.2).

The quality of the underlying vowel may be taken initially to be low-mid *e, o*. The former height of the nuclear vowel is irrelevant, that is, height is irrelevant for the new nasals. The front/back opposition is reinforced within the system, as also is the feature [Round].

Diphthongs ending in a liquid sonorant (*r, l*) This set of diphthongs is reflected in various distinct ways over the whole group, indicating that it had its final realization after the beginning of the break-up of the group. The motive force of syllable opening belongs to the unified period (before the sixth century, since the period of break-up is roughly sixth to ninth century), and so we treat it in this chapter. Its relative lateness may be related to the phonetic nature of the liquid sonorants, which are particularly able to function as vocalic nuclei themselves – witness the many languages, especially Slavic ones, in which both /r/ and /l/, or at least /r/, may be nuclei: Czech, Slovak, B/C/S, Slovenian, Macedonian.

The process here is not always strictly "monophthongization", but, rather, restructuring of the diphthong. In this set, the height of the nuclear vowel *is* relevant. Perhaps the lateness of the results means that the new vowel system was well in place by then, so that the different height of the former high vowels – that is, the height difference between the new /u/, /i/, on the one hand, and /ъ/, /ь/, on the other – is well established. This means that these latter vowels would not be fused with the low /e/, /o/ as happened with the nasals. Or perhaps it is simply that the combination of high(er) and/or short(er) /ъ/ and /ь/ with / r/, /l/ led more easily to a syllabic sonorant.

Thus we see in principle different results for ъ/ь + *r*/*l* and *o*/*e* + *r*/*l*. But in addition we see different results for each set across the dialectal spectrum. The most common way of formulating this structure is to use 'C' (some use 't') for any consonant and 'R' for *r*/*l*. So we are dealing with the late Proto-Slavic structures *CoRC* and *CeRC* (7c), *#oRC* (that is, where /o/ is word-initial) (7d), and *CъRC*, *CьRC* (7e). In some cases *r* and *l* behave differently. We treat the

forms with /o/ and /e/ first, since they all result in an open syllable, and are thus relatively early. The forms with /ъ/ and /ь/ may not have reached the open-syllable stage throughout the area, and so may be relatively late.

CoRC, CeRC For this structure we have three distinct groups, though initially there may only have been two types. There are really only two possibilities for resolution of the problem where the nuclear vowel is not of the reduced sort: (1) insertion of a new vowel (epenthesis), thereby creating an extra syllable ('type 1'); or (2) metathesis (inversion) of the vowel/sonorant sequence ('type 2'). Solution (1) is realized in East Slavic, and solution (2) in the Lekhitic area (northwest: Polish and Sorbian). A third type ('type 3') is a variant of metathesis, in which the vowel has additionally been lengthened ($o > a$, $e > ě$), and is found in the 'southern' area (South Slavic, Czech and Slovak, see 1.4.1). A common interpretation of type 2 is that it actually began as type 1, with epenthesis. Then the original vowel was lost and the epenthetic one remained, an analysis for which there is some circumstantial evidence. It is conceivable that type 3 also began as epenthesis, rather than metathesis, but there is no attested evidence of this. The three types, based on the modern situation, are shown in (7c):

(7c) restructuring of diphthongs of '*o*/*e* + liquid sonorant'

a	East:	CoRC (?via CoRəC) > CoRoC; CeRC (?via CeRɪC) > CeReC
b	Pol/Sorb:	CorC (?via CoRəC > CoRoC) > CRoC CeRC (?via CeRɪC > CeReC) > CReC
c	South + Cz/Slk ('southern'):	CoRC (?via CoRəC > CoRoC) > CRōC > CRaC CeRC (?via CeRɪC > CeReC) > CRēC > CRěC

(Late) PSl		B/C/S	Rus	Cz	Pol
**kórva*	'cow'	*krȁva*	*koróva*	*kráva*	*krowa*
**bērgъ*	'bank'	*brêg*	*béreg*	*břeh*	*brzeg*
**zôlto*	'gold'	*zlâto*	*zóloto*	*zlato*	*złoto*
**mél-ti*	'grind'	*mlȅ-ti*	*molót'*	*mlét*	*mleć*
**žēlbъ*	'gutter'	*žlēb*	*žólob*	*žleb*	*żłób*
**želzá*	'gland'	*žlézda*	*železá*	*žláza*	*zołza*

The Eastern result is referred to in Russian as "*polnoglasie*", in English as "pleophony" or "full vocalization". As suggested in the bracketed forms, it is

possible that the Lekhitic group went through this step also, but then removed the first vowel. The point-to-point statement could equally well be simply metathesis, as is accepted to be the case for the south, which then lengthened the one vowel, possibly in compensation.

At the suprasegmental level, the above picture is somewhat more complicated, since the stress in East Slavic may land on either the old or the new (inserted) vowel, and in the other languages the new vowel may be accompanied by different quantity or pitch. The cause of these variations is the nature of the pitch (rising or not) on the original diphthong (as marked on the examples in (7c) by ´, ˆ respectively).

#oRC In theory we should also find the structure *#erC*. But there are no reliable examples, so it is normally excluded from consideration, though of course one can say how it might have developed. We have here two results, both described as metathesis, one with lengthening of the vowel in some forms. The inserted-vowel approach is, in theory, a possible intermediate step, as suggested for the Lekhitic group (type 2) above. But in this case there is no secondary evidence which might support it. Moreover, the isoglosses are different, since all of East and West Slavic have the same result – simple metathesis in some forms, metathesis plus lengthening in others – while the South always has metathesis plus lengthening. Thus, the South is consistent in its reflexes of the initial and medial contexts, but the East and the West are not. It is presumed that the initial position presented other factors, which caused either an earlier or later shift. Such factors include the vulnerablility of the absolute initial vowel position in a language shifting towards open syllables, since the preceding sound, that is, the end of the preceding word, would now be ending in a vowel. This produced undesirable hiatus (e.g. *ne orvьn-*'not (un-)even'). In other contexts the typical solution taken by Proto-Slavic was to insert a prothetic glide (*u̯* or *i̯*) which later became a consonant (*w*/*v* or *j*) (see further 3.2.1.4). In this particular context metathesis may have been seen as a more attractive solution:

(7d) restructuring of diphthongs of 'initial *o* + liquid sonorant'
a East/West: oRC > RoC/RōC > RoC/RaC
b South: oRC > RōC > RaC

(Late) PSl	B/C/S	Rus	Cz	Pol
**ôrvьn-* 'even'	*rȃv(a)n-*	*róv(e)n-*	*rovn-*	*równ-*
**órdlo* 'plough'	*rȁlo*	*rálo*	*rádlo*	*radło*
**ôlkotь* 'elbow'	*lȃk(a)t*	*lók(o)t′*	*lok(e)t*	*łok(ie)ć*
**ólk-om-* 'hungry'	*lȁkom-*	*lákom-*	*lakom-*	*łakom-*

The two results in East/West are again caused by pitch differences in the underlying diphthong. A rising pitch was responsible for the lengthened form ($> \bar{o} > a$) of the new vowel.

As for the potential *#eRC*, it "should" therefore have become *ReC*/*RěC* in the East and West, and *RěC* in the South. One frequently produced example matching the criteria is **érdъk-* 'rare': B/C/S *réd*(*a*)*k*-/*rijèd*(*a*)*k*-, Rus *réd*(*o*)*k*-, Cz *řídk*-, Pol *rzadk*-; this etymology, based on the Lithuanian form *erdvas* 'spacious' (alongside *retas* 'rare'), is, however, challenged by Vasmer (1964–1973, vol. 3), for whom the Proto-Slavic form is **rědъk-*.

CъRC, CьRC There are two views on the behavior of these diphthongs. One view (e.g. Schenker, 1993) holds that they were the first to become restructured with the opening of the syllable, facilitated by the ability of the liquids to function as vocalic nuclei, which allowed the conversion of the group to syllabic /r̥/ and /l̥/. The second (e.g. Carlton, 1990) claims that they were actually the last to change, as evidenced by their failure to open in some areas. The second view has the problem of explaining why just this type failed to open, especially given the inherent vocality of the liquids. The first has the problem of accepting a see-sawing effect of syllabicity. But this is not difficult, since it is clear that there were several episodes of this effect between PIE and modern Slavic. However, it has the more important problem of accepting that the syllabic segments retained the opposition of hard/soft (±Palatalized) in some form.

The tendency for the two *jers* (the name given to the vowels ъ and ь from the old name for these Cyrillic letters) to be reduced, and in many areas to end up as *schwa* ([ə]) (1.3.1.6), would easily have allowed these sequences to become syllabic sonorants. In order to account for the results, we must assume that the back/front quality of the vowel was at least initially preserved in the hard/soft varieties of the new syllabic sonorant. In this view, the syllabic sonorants are early Proto-Slavic, while the second stage, after the break-up, involves their retention in some areas, and their restructuring – once again – as sequences of 'vowel + liquid'. In this restructuring, the vowel is often "correctly" reconstituted as the back or front reflex of the *jer*, which can only occur if there remained a distinction in the syllabic versions. So long as one accepts this possibility, this view has the merit of conforming to the other structural developments of Proto-Slavic. But the view which accounts for the "correct" vowel reflexes through the failure to develop syllabic versions has the structural problem of accounting for the retention of closed syllables. We prefer the first view, and so shall assume the development of hard/soft syllabic liquids in Proto-Slavic and offer in (7e) only Proto-Slavic examples (roots), leaving the details of their further development to chapter 3 (3.2.1):

(7e) restructuring of diphthongs of *jer* + liquid sonorant in Proto-Slavic

CъrC > Cr̥C: *tъrg-* 'trade' > *tr̥g-*; *gъrdl-* 'throat' > *gr̥dl-*

CъrC > Cr̥′C: *pъrv-* 'first' > *pr̥'v-*; *vъrxъ* 'summit' > *vr̥'x-*

(**k*ъrn- >) *čъrn-* 'black' > *čr̥'n-*

CъlC > Cl̥C: *dъlg-* 'debt' > *dl̥g-*; *tъlstъ(jь)* 'fat' > *tl̥st-*

CьlC > Cl̥′C: *dьlg-* 'long' > *dl̥'g-*; *vьlkъ* 'wolf' > *vl̥'k-*

(**g*ьlt- >) *žьlt-* 'yellow' > *žl̥'t-*

In all the above cases of diphthongs, wherever the diphthong was followed by a vowel, the solution to the open syllable impetus was simply to shift the new boundary to after the nuclear vowel. The former semivowel or sonorant became a syllable-initial consonant: *i̯* > *j*, *u̯* > *w* (> *v*), *m, n, r, l*. This is an important source of often quite complex morphophonological alternations, based on whether inflections began with a vowel or a consonant (4.4).

1.3.1.8 *Suprasegmental*

The following late changes occurred affecting all areas: (1) tone became restricted to stressed position; (2) all vowels which had rising pitch (automatic on the old long pure vowels, phonemic on those derived from diphthongs) were shortened, meaning that these vowels now had phonemic quantity under stress, since it was no longer predictable from quality. Furthermore, inasmuch as the new short vowels retained their rising pitch, tone was no longer limited to long vowels (and former diphthongs); (3) final long vowels were shortened, but pretonic length was preserved. Both quantity and tone were thus phonemic in a wide range of vowels, the latter only in stressed position.

1.3.1.9 *The Late Proto-Slavic phonological system*

We may now consider the system which has arisen by the end of the common period of development, just before the break-up in the sixth century, where we see different results for jointly motivated changes and the start of locally motivated changes. Table 1.3 may be compared with that of early Proto-Slavic (table 1.1). Brackets indicate regional or temporary variants. The combination of 'labial + *l*' (as the 'jotated' version) probably did not arise in the West. Of the soft dentals (from PV2/3) only /c'/ (from /k/) was general. The palatal stops are conveniently described as still general in this form throughout the area; /g/ and /v/ have alternative articulations by region, as stop *vs* fricative for the first, as labio-dental *vs* bilabial for the second. The apostrophe is used to distinguish palatal sounds from simply palatalized ones (marked with acute).

Table 1.3. *Late Proto-Slavic phonological system*

Consonants				
	Labial	Dental	Palatal	Velar
Stop	p (pl')	t	t'	k
	b (bl')	d	d'	(g)
Nasal	m (ml')	n	n'	
Fricative		s (s′)	š'	x
	(v) (vl') (w)	z (z′)	ž'	(ɣ)
		j		
Affricate		c′	č'	
		(dz′)	(dž')	
Liquid		r	r'	
		l	l'	
Vowels				
	Front	Central	Back	
High	i	y	u	
High-mid	ь		ъ	
Low-mid	e ę		o ǫ	
Low	ě		a	
Syllabic Sonorants				
	r̥, r̥′ l̥, l̥′			
Suprasegmental				
Stress	Free			
Quantity	On many vowels under stress or in pretonic position			
Tone	On many vowels under stress			

Note: there were also three fronted allophones after palatal consonants: ü, ǫ̈, ä.

1.3.2 Morphology

1.3.2.1 *Nominal*

The following are the features and categories of the late PIE nominal system and how they were treated in Proto-Slavic (discussion of the actual forms will be taken up in later chapters and sections as appropriate, especially chapter 5):

Case: the seven cases of PIE – nominative, accusative, genitive, dative, instrumental, locative, ablative, plus the vocative – were reduced to six plus vocative in Proto-Slavic (and Baltic) by the conflation of the ablative and genitive (into genitive);

Number: of the three numbers – singular, dual, plural – the dual was already losing ground in Proto-Slavic, having its range of cases reduced to three by the syncretization of NOMINATIVE + ACCUSATIVE, GENITIVE + LOCATIVE

and DATIVE + INSTRUMENTAL. Subsequently it was completely lost except in Slovenian and Sorbian, but it was still functioning in at least OCS and Russian Church Slavonic in the "Old Russian" period (2.3.1, 5.4.1);

Gender: the three genders (masculine, feminine, neuter) were retained, inherent (syntactic) in substantives and agreeing (morphological) in adjectives (including participles), pronouns and some numerals (1–4);

Proto-Slavic refined the masculine group with sub-categories of ±Personal and ±Animate (5.4.4);

Adjectives: no change in gradation – positive, comparative and superlative; an added feature of ±Definite in most (non-possessive) adjectives (5.4.3);

Pronouns: no change to the general range and type.

1.3.2.2 *Verbal*

Tense: the six tenses of late PIE (present, future, aorist, imperfect, perfect, pluperfect) were all retained in Proto-Slavic. But their formation in many cases was different: the future, perfect and pluperfect were re-formed analytically with auxiliary verbs and either the infinitive or a past participle;

Mood: the four-way system of PIE (indicative, subjunctive, optative, imperative) became a three-way one in Proto-Slavic, with the replacement of the imperative forms by those of the optative and the functional loss of the latter; the subjunctive forms became primarily conditional, again analytic in form;

Voice: PIE active and middle were redefined in Proto-Slavic as ±Reflexive, and a new passive was added, a participial form like that of English *-en*;

Aspect: the most important development in Proto-Slavic was the gradual shift in the relative importance of aspect over tense. The feature ±Complete (perfective/imperfective), while already present in PIE's past tenses, became more important than that of time. A further development was the splitting of imperfective motion verbs into ±Determinate (or ±Continuous).

Person: no difference, the three persons – 1st, 2nd, 3rd – were retained.

Two new, non-finite, verbal categories in Proto-Slavic were the infinitive and the supine, both derived from PIE deverbal nouns with an added *-t* suffix and an inflexion, frozen into indeclinable forms.

1.3.3 Syntax

Little more can be said about the details of syntax of either PIE or Proto-Slavic than follows from the morphological changes noted above. Given the continued

high degree of inflection in Proto-Slavic, little would have had to change in terms of word order, and the overall categories able to appear in a given syntactic position would likewise have been the same in principle.

1.4 The sub-division of Slavic

The standard classification of Slavic involves a three-way grouping:

a. East Slavic: Russia, Ukraine and Belarus, including Siberia and the Far East through the extension of Russian into Asia;
b. West Slavic: to the west of East Slavic within northern Europe;
c. South Slavic: the Balkans, from Slovenia south and east to Macedonia and Bulgaria.

These three groups reflect the three major dialects of Slavic after the break-up of Proto-Slavic unity. The modern Slavic languages then further sub-divided from these three main groups. However, there are numerous features which cut across this underlying classification between individual languages, or, in some cases, groups of languages. The future South Slavs would have entered the Balkan Peninsula via both the west and east of the Carpathian Mountains. Those to the west (the future Serbs, Croats and Slovenes) came from the earlier "south-west Slavic" group, while those to the east (the Bulgarians and Macedonians), followed the same way later by the non-Slavic Bulgarians, were from the "south-east Slavic" group. From this scenario follows the linguistic grouping observed in modern times, including parallels between, say, Slovak and Slovenian/Croatian, on the one hand, and Russian and Bulgarian, on the other (see chapter 10).

In the sections which follow, we present a hierarchical discussion of the characteristics of the Slavic languages. We begin with the macro-features which define the three major groups (which we shall call 'Stage 1 features'), then with sub-groups within the three major groups ('Stage 2 features') and, finally, with the individual languages ('Stage 3 features') (1.5–1.7). Unless otherwise specified, any language can be taken to have not only the Stage 3 features listed with it, but also the Stage 1 and Stage 2 features which belong further up its tree. Polish, for instance, also shares the features for Lekhitic (2) and West Slavic (1). We include not only features unique to a language group or language but also features which are commonly regarded as being among the typical characteristics: for instance, *akan'e* (the loss of distinction of low and low-mid vowels, especially /o/ and /a/), is found in both Russian and Belarusian (3.2.1.5).

Most of the features which we discuss involve phonology, morphophonology and morphology, together with some broad-scale lexical features. This reflects not

only the traditional orientation of Slavic genetic linguistics but also the fact that the syntax of Slavic is less differentiated than these other levels of language. To facilitate comparison and consultation, the data are arranged in this order: phonology (vocalic system, consonantal system, suprasegmentals); morphophonology; morphology; syntax; lexis. This order also reflects the order of the chapters in this book. As far as possible in the following we avoid repetition of examples by cross-referring to other locations of relevant discussion and examples.

1.5 **South Slavic**

South Slavic covers Slovenian, B/C/S (including Serbo-Croatian, see 2.2.4), Macedonian and Bulgarian. Old Church Slavonic, the earliest Slavic liturgical and written language, also belongs to South Slavic (2.2.1). The South Slavs inhabit a geographically coherent area, and are now separated from West and East Slavic by Austria, Hungary and Romania.

1.5.1 Stage 1 features of South Slavic

The phonological grouping of South Slavic is more clear-cut than its morphological (chapter 5) or syntactic (chapter 7) grouping, where the eastern South Slavic languages (Bulgarian and Macedonian) are distinct not only from the other South Slavic languages but also from other non-South Slavic languages. However, there are few features which have united the whole group in the past, and even fewer in modern times:

1. The strong *jers* probably merged into *schwa* ([ə]), but this later shifted to [a] (B/C/S) or split according to the retained hard/soft quality of the preceding consonant into /o e/ (Mac), /ə e/ (Blg) (3.2.1.1, table 3.1)
2. The front nasal *ę* > /e/ (3.2.1.2, table 3.2)
3. Loss of /y/, fused with and into /i/ (3.2.1.5, 3.3.1) (PSl *syn-* 'son' *vs sin-* 'blue': Rus *syn-, sin′-*; all South Slavic has *sin-* for both)
4. Syllabic liquids were initially retained, but /l̥/ was then lost everywhere (> 'vowel + /l/' or '/l/ + vowel' in Bulgarian, Macedonian and Slovenian, > /u/ in B/C/S) and /r̥/ became [ər/rə] in Bulgarian. A lost *jer* after a liquid produced the same results (3.2.1.1, 3.2.2.6 (42a–b))
5. Hardening of palatals and dental affricates (that is, lowering of the tongue blade), e.g. š' > š, č' > č, c′ > c.

6. *tl dl* > *l*
7. *CoRC* etc. > *CRaC* (1.3.1.7 (7c–d) above)

1.5.2 Stage 2 features of South Slavic

The major sub-groups are "western": B/C/S and Slovenian, and "eastern": Bulgarian and Macedonian (and OCS). They derived from different origins and routes into the Balkans and/or different contacts once in the Balkans, for example the closer contacts of the eastern group with the Turks (the Bulgarians and Macedonians were under the domination of the Ottoman Empire from 1396 to 1897). This may account for their greater "Balkanization" – see the morphological features marked "Balkan" below.

1.5.2.1 *Features in support of the traditional Stage 2 grouping*

a. Phonology

1. Initial *je-* > *e* in the east, not in the west (3.2.1.4 (7)):

   (8) 'lake': B/C/S *jèzero* Sln *jézero* Blg, Mac *ézero*

2. Strong *ь* > *e* in the east, not the west (3.2.1.1, table 3.1)
3. Vocalic quantity/tone in the west, not the east (3.5)

b. Morphology and word formation

Eastern:

1. The loss of most of the case forms of the declensions (Balkan) (5.4.2, 5.5.1)
2. The 3 Person pronoun is *toj* 'he', related to the masculine singular nominative of 'that' (B *tója*, M *toj*), where the rest of Slavic has forms related to *on* (5.5.2.3)
3. Post-posed definite article (Balkan) (5.4.3, 5.5.1, 8.1.1)
4. The 3 Person plural present of all verbs ends in *-at* (the reflex of the back nasal -**ǫ*-*t*); in the west this ending is *-e-* class only) (5.5.5.4)
5. The loss of the infinitive, which is replaced by subordinate clauses introduced by *da* 'in order to' (Balkan). Serbian has both, but prefers the latter, while Croatian is the reverse (5.5.5.3)
6. Bulgarian and Macedonian have taken over from Turkish the concept of "renarrative" verb forms, in which a separate set of inflexional forms of the verb are used to mark events which the speaker is unable to vouch for as fact. This has resulted in totally new paradigms of renarrative verb forms (Balkan) (5.5.5.8).