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*.

1

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 funda-mental 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

4   *0. Introduction*

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 wide-spread 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*
	dràga		*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.