20 Indo-Aryan Languages

George Cardona

1 Introduction

Indo-Aryan languages, the easternmost group within Indo-European, are spoken by approximately five hundred million persons in India, Pakistan, Bangladesh, Nepal and other parts of the Himalayan region, as well as in Sri Lanka. Gypsy (Romany) dialects of the USSR, the Middle East and North America are also of Indo-Aryan origin. Indo-Aryan is most closely related to Iranian, with which it forms the Indo-Iranian subgroup, speakers of which shared linguistic and cultural features, including a name they called themselves (Sanskrit *ārya-*, Avestan *airya-*). Among the innovations that characterise Indo-Iranian is the merger of Proto-Indo-European *ĕ*, *ŏ*, *ă* into *ă*: Skt. *asti* 'is', *pati-* 'master, husband', *ajati* 'leads', *dadhāti* 'puts, makes', *dadāti* 'gives', *mātr̥-* 'mother': Av. *asti*, *paiti-*, *azaiti*, *dadāiti* ('puts, makes, gives'), *mātar-*: Gk. *estì*, *pósis*, *ágei*, *títhēsi*, *dídōsi*, *mắtēr* (Dor.). Two major phonological features distinguish Indo-Aryan from the rest of Indo-European, including Iranian. One of these is an inherited property: Indo-Aryan retains voiced aspirated stops, as in Skt. *gharma-* 'warmth', *dadhāti*, *bharati* 'carries'. The other is an innovation: Indo-Aryan languages distinguish dental and retroflex stops. Originally, retroflex *-ḍ-*, *-ḍh-* arose through sound changes, as in Skt. *nīḍa-* 'resting place, nest', *mīḍha-* 'reward', with *-īḍ-*, *-īḍh-* from *-iẓḍ-*, *-iẓḍh-* (< *-izd-*, *-izdh-*). Such developments resulted in contrastive retroflex stops, albeit restricted, and the compass of such consonants was extended through borrowings from Dravidian languages. Most Indo-Aryan languages still have voiced aspirates and retroflex stops, although in certain ones, abutting on non-Indo-Aryan languages, these contrasts have been reduced: Sinhalese (Sinhala) has no aspirated stops, Kashmiri lacks voiced aspirates and Assamese (Asamiya) has no retroflex stops.

Old Indo-Aryan is represented in numerous sources (see the chapter on Sanskrit). The earliest preserved Middle Indo-Aryan documents are Aśoka's edicts (third century BC), in various dialects. Middle Indo-Aryan languages were also used for other literary, philosophical and religious works. The Buddhist canon and later treatises of Theravada Buddhism are

in Pāli, the Jaina canon in Ardhamāgadhī; Jainas also used Jaina Māhārāṣṭrī and Śaurasenī in works. The literary exemplar of Middle Indo-Aryan, however, is Māhārāṣṭrī, and the most advanced stages of Middle Indo-Aryan developments are found in Apabhraṁśa dialects, used as literary vehicles from before the sixth century. All Middle Indo-Aryan varieties can be subsumed under the label Prakrit (Skt. *prākr̥ta-*, Pkt. *pāia-* 'stemming from the original, natural'), referring to vernaculars in contrast to the polished language called *saṁskr̥ta*. Traditionally, most Indian commentators and grammarians of Prakrits derive these from Sanskrit, but there are formations in Prakrits found in Vedic sources but not in Classical Sanskrit. Thus, as Classical Sanskrit is not derivable from a single attested Vedic dialect, so the Prakrits cannot be derived from Classical Sanskrit. In the present sketch, I use *Prakrit* in a narrow sense, of Middle Indo-Aryan languages other than Aśokan dialects, Pāli or Apabhraṁśa. There are abundant literary sources for New Indo-Aryan languages from the twelfth century on, some materials from earlier times.

Several scripts have been and currently are used for Indo-Aryan languages. In ancient times, two major scripts were used on the subcontinent: Kharoṣṭhī, written from right to left, was predominantly used in the north-west, Brāhmī, written from left to right, elsewhere. Most scripts used for Indo-Aryan languages stem from Brāhmī, including Devanāgarī (see section 2 of the chapter on Sanskrit), widely employed for Sanskrit and now the official script for Hindi, Marathi, Nepali. The Arabic script, with modifications, is used for some Indo-Aryan languages, including Urdu.

2 Phonological and Grammatical Developments

In the following, I sketch major phonological and grammatical developments that characterise Middle and New Indo-Aryan, using Old Indo-Aryan as a point of reference (see sections 1.2, 2 of the chapter on Sanskrit).

2.1 Phonology

In Middle Indo-Aryan, word-final consonants other than *-m*, which developed to *-ṁ* with shortening of a preceding vowel, were lost: Skt. *putrāt* (abl. sg.) 'son', *putrās* (nom. pl.), *putram* (acc. sg.): Pāli *puttā*, *puttaṁ*. Interior clusters of dissimilar consonants were generally eliminated through assimilation (as in *puttā*) or epenthesis: Skt. *sakthi-* 'thigh', *varga-* 'group', *agni-* 'fire', *śukla-* 'white', *pakva-* 'cooked, ripe', *satya-* 'true', *adya* 'today': Pāli *satthi-*, *vagga-*, with assimilation of the first consonant to the second, *aggi-*, *sukka-*, *pakka-*, with the second consonant assimilated to the first, and *sacca-*, *ajja-*, with palatalisation; similarly, Skt. *rājnā* (inst. sg.) 'king', *rājñas* (gen. sg.): *rāññā*, *rāñño* in the Girnār version of Aśoka's first rock edict, but *lājinā*, *lājine*, with epenthesis, in the Jaugaḍa version. Generally, a nasal

remains unassimilated before an obstruent: Skt. Pāli *danta-* 'tooth'. Metathesis applies in clusters of *h* with nasals or *y*, *v*: Skt. *cihna-* 'mark', *sahya-* 'to be endured', *jihvā-* 'tongue': Pāli *cinha-*, *sayha-*, *jivhā-*. Clusters of voiceless spirants with obstruents develop to obstruent sequences with aspiration: Skt. *paścāt* 'afterwards', *hasta-* 'hand': Pāli *pacchā*, *hattha-*. Further, clusters with voiceless spirants and nasals show voice assimilation and metathesis, resulting in nasals followed by *h*: Skt. *tr̥ṣṇā-* 'thirst, longing': Pāli *taṇhā-*. Initial clusters changed in the same ways, with subsequent simplification: Skt. *prathama-* 'first', *tyajati* 'abandons', *skandha-* 'shoulder', *snāti* 'bathes': Pāli *paṭhama-*, *cajati*, *khandha-*, *nhāyati*. In compounds and preverb-verb combinations where the assimilated cluster was intervocalic, it was retained, resulting in alternations such as Pāli *pamāṇa-* 'measure': *appamāṇa-* 'without measure, endless' (Skt. *pramāṇa-*, *apramāṇa-*). In early Middle Indo-Aryan, word-internal single consonants were retained, as shown in examples cited. Later, as exemplified in Māhārāṣṭrī, non-labial non-retroflex unaspirated obstruents were generally deleted, and *p*, *b* changed to *v*: *loa-* 'world, people', *naa-* 'mountain', *paura-* 'ample', *gaa-* 'elephant', *viāṇa-* 'awning', *savaha-* 'oath': Skt. *loka-*, *naga-*, *pracura-*, *gaja-*, *vitāna-*, *śapatha-*. Presumably, an intermediate step prior to loss involved the voicing of consonants, and some dialects reflect this; for example, in Śaurasenī intervocalic dentals were voiced (*ido* 'hence', *tadhā* 'thus': Skt. *itas*, *tathā*), and *thūbe* 'stupa' (Skt. *stūpas*) occurs in Aśokan. The loss of consonants resulted in word-internal sequences of vowels that were not found in Old Indo-Aryan, though such vowels were separated by *y*, *v* in some dialects. Intervocalic non-retroflex aspirates generally changed to *h*, but *-ṭ-*, *-ṭh-* were voiced, and *-ḍ-* developed to *-ḷ-*, whence *-l-*: Pkt. *sāhā-* 'branch', *meha-* 'cloud', *naḍa-* 'actor', *maḍha-* 'cloister' (Skt. *śākhā-*, *megha-*, *naṭa-*, *maṭha-*), Skt. *krīḍati* 'plays': Pāli *kīḷati*, Pkt. *kīlai*. The spirantal system of Old Indo-Aryan was also generally simplified. On the evidence of Aśokan documents, dialects of the extreme north-west retained *ś ṣ s*, as in Shāhbāzgaṛhī *paśucikisa* 'medical treatment for cattle', *vaṣeṣu* (loc. pl.) 'years'. But elsewhere the sibilants merged to *s*, and later in the east, as represented by Māgadhī, one has *ś* (e.g. *keśeśu* (loc. pl.) 'hair', *śahaśśa-* 'thousand': Skt. *keśeṣu*, *sahasra-*). In Apabhraṁśa, *-s(s)-* developed to *-h-*, as in *taho* 'of that' (Pāli *tassa*, Skt. *tasya*), and intervocalic nasals lost their occlusion, resulting in nasalisation, as in *gāũ* 'village' (Pkt. *gāmo*, Skt. *grāmas*), *pasāẽ* 'through the grace of' (Pkt. *pasāeṇa*, Skt. *prasādena*).

The Middle Indo-Aryan vowel system also shows major developments. As shown, word-internal vowel sequences not permitted earlier now occurred. Conversely, overheavy syllables — with long vowels followed by consonant clusters — permissible in Old Indo-Aryan, were eliminated, through shortening of vowels or reduction of clusters. Moreover, as -V̄C- and -V̄CC- were prosodically equivalent, one has either as reflex of earlier -V̄C-, -V̆CC-. For example: Skt. *lākṣā-* 'lac', *dīrgha-* 'long', *śvaśrū-* 'mother-in-law', *sarṣapa-* 'mustard seed': Pāli *lākha-*, *dīgha-*, *sassū-*, *sāsapa-*: Pkt. *lakkhā-*, *diggha-*/*dīgha-*, *sāsū-*, *sāsava-*. In addition, vocalic *r̥* is replaced by various vowels; *ai*, *au*, were monophthongised to *e*, *o*; *-aya-*, *-ava-* developed to *-e-*, *-o-*; and short *ĕ*, *ŏ* arose through shortening before clusters: Skt. *r̥kṣa-* 'bear', *vr̥ścika-* 'scorpion', *pr̥cchati* 'asks', *taila-* 'oil', *jayati* 'is victorious', *prekṣate* 'looks', *aurasa-* 'legitimate', *bhavati* 'is', *maulya-* 'price': Pāli *accha-*, *vicchika-*, *pucchati*, *tela*, *jeti*, *pekkhati*, *orasa-*, *hoti*, *molla-*. Moreover, many of the complex morphophonemic alternations that applied in Old Indo-Aryan across word boundaries (see section 1.2 of the chapter on Sanskrit) were eliminated. Certain phonological developments also characterised major dialect areas. As noted, the extreme north-west retained different sibilants. In addition, at Aśoka's time the extreme west and east respectively were characterised by having *r*, consonant assimilation and *-o* for earlier *-as* and its variants as opposed to *l*, a tendency to epenthesis and *-e*: *rāñño* versus *lājine*.

Some of the tendencies observed earlier continue in evidence into New Indo-Aryan. Thus, the resolution of -V̆CC- to -V̄C- takes place in some areas: Gujarati *pākũ* 'ripe', *lāḍu* 'a sweet': Hindi *pakkā*, *laḍḍu*. Though *ai*, *au* are retained well into the modern period and still found, they are also monophthongised, as in Hindi *hɛ* 'is', *cɔthā* 'fourth' (spelled *hai*, *cauthā*). Middle Indo-Aryan *ḍ*, *ḍh* develop to flaps (but the etymological spellings are retained) except in initial position and after ṇasals; e.g., Hindi *sāṛī* 'sari' (Pkt. *sāḍiā-*). In the north-west, assimilation affects a sequence of a nasal with an obstruent: Panjabi *dand* 'tooth' versus Hindi *dā̃t*. On the other hand, the widespread loss of earlier final vowels results in word-final consonants, although in certain areas the final vowels are retained; e.g. Panjabi *dand*, Hindi *dā̃t*, but Sindhi *Dandu*. The last has an initial imploded stop, characteristic of Sindhi and some adjacent languages. Dialectal developments have resulted in other phonological features not found in Middle Indo-Aryan. For example, Panjabi developed a tonal system; Kashmiri has developed pharyngealised consonants; in languages of the south-west there are two sets of affricates, as in Marathi *c* (= *ts*) versus *č*; and languages of the extreme east have rounded the vowel *a*, as in Bengali (Bangla), where one also finds limited vowel harmony.

2.2 Morphology and Syntax

The grammatical system of Middle Indo-Aryan is characterised by a general reduction of complexities in comparison with Old Indo-Aryan. The dual is eliminated as a category distinct from the plural. The trend to replace variable consonant stems with single stems ending in vowels, already evident in Old Indo-Aryan (e.g. Skt. *danta-* 'tooth', earlier *dant-*/*dat-*), continues: Pāli *gacchanta-* 'going' (masc. nom. sg. *gacchanto*, gen. pl. *gacchantānaṁ*) as against Skt. *gacchant-*/*gacchat-* (see section 2.2.2 of the chapter on Sanskrit). The loss of final consonants also contributed to the steady

elimination of consonant stems, e.g. Pāli *āpā-* 'emergency', *sappi-* 'butter': Skt. *āpad-*, *sarpis-*. The nominal case system too is reduced. At an early stage, the dative is replaced by the genitive except in expressing a goal or purpose: Pāli *etesaṁ pi abhayam dammi* 'I grant (*dammi*) them too (*etesaṁ pi*) security' has a genitive *etesaṁ* construed with *dammi*, and Jaina Māhārāṣṭrī *namo tāṇaṁ purisaṇaṁ* 'homage to those men' has a genitive in construction with *namo*. Formal datives occur in examples like Aśokan *etāya atthāya idaṁ lekhāpitaṁ* 'this (*idaṁ*) has been caused to be written (*lekhapitaṁ*) for this purpose (*etāya atthāya*)', Pāli *jhassu rūpaṁ apunabbhavanāya* 'give up (*jhassu*) your body (*rūpaṁ*) so as not to be born again (*apunabbhavanāya*)'. In addition, nominal and pronominal types are less strictly segregated, as can be seen from *etāya*, *tāṇaṁ* (Skt. *etasmai*, *teṣām*) in examples cited.

Although early Middle Indo-Aryan retains middle forms, the contrast between active and medio-passive in the verb system is generally obliterated. Thus, Pāli has *maññati* 'thinks', *jāyati* 'is born' and passives of the type *vuccati* 'is said', with etymologically active endings; contrast Skt. *manyate*, *jāyate*, *ucyate*. The contrast between two kinds of future formations is absent in Middle Indo-Aryan, which has the type Pāli *hossati* 'will be'. Further, the distinction among aorist, imperfect and perfect is obliterated. With few exceptions, the sigmatic aorist supplies the productive preterit. Thus, Pāli has several preterital formations, but the productive one is sigmatic and based on the present stem, not on the root as in Old Indo-Aryan: *ahosi* 'was' (3 sg.), *ahosuṁ* 'were' (pres. *hoti honti*), *agacchi*, *agacchisuṁ* (*gacchati*, *gacchanti*). In later Middle Indo-Aryan, verbally inflected preterits are generally given up in favour of participial forms, as in Śaurasenī *mahārāo vi āado* 'the king (*mahārāo*) also (*vi*) has arrived (*āado*)', where *āado* agrees in case, number and gender with *mahārāo*. The participle of a verb that takes a direct object shows object agreement: in Jaina Māhārāṣṭrī *teṇa vi savvaṁ siṭṭhaṁ* 'he too has told everything', *teṇa* (inst. sg.) refers to the agent, and *siṭṭhaṁ* 'told' agrees with *savvaṁ* (nom. sg. nt.) 'everything'. If no object is explicitly referred to, the neuter nominative singular of a participle is used; e.g., Jaina Māhārāṣṭrī *pacchā raṇṇā cintiyaṁ* 'afterwards, the king (inst. sg. *raṇṇā*) thought (*cintiyaṁ*)'.

Alternations of the type Skt. *asti–santi* (see section 2.2.3 of the chapter on Sanskrit) are eliminated in Middle Indo-Aryan, where the predominant present formation involves a single stem: Pāli *eti* 'goes' *enti* 'go', *sakkoti–sakkonti* (*sak* 'be able'), *chindati–chindanti* (*chid* 'cut'). Stems like *chinda-* reflect a generalisation, based on a reanalysis of third plural forms, of stems with *-a*. The elimination of strictly athematic presents with variable stems allowed the use of the second singular imperative *-hi* in a domain wider than this had in Old Indo-Aryan; e.g., Pāli *jīvāhi* 'live' (Skt. *jīva*). Similarly, optatives with *-e-* and *-yā-* are not sharply segregated; a form like Pāli *bhaveyya* (3 sg.) shows a blend of the two. Middle Indo-Aryan

continues to use morphological causatives with *-i-/-e-* (Pāli 3 sg. pres. *kāreti*), but the type in *-āpe-* (Pkt. *-āve-*) is extended beyond its earlier domain, as in Pāli *vasāpeti* 'has ... stay'.

Nominal forms of the Middle Indo-Aryan verb system are of the same types as in Old Indo-Aryan: present and past participles (see above), gerundives (Pāli *kātabba-* 'to be done', *dassanīya-* 'worthy of being seen'), gerunds, infinitives, with some innovations. For example, Pāli *nikkhamitvā* 'after leaving' has *-tvā-* after a compound, and *pappotuṁ* has *-tuṁ* added to the present stem, not the root. Contrast Skt. *niṣkramya*, *prāptum*.

The late Middle Indo-Aryan stage represented in Apabhraṁśa foreshadows New Indo-Aryan in several ways. Forms of the nominal system with *-au*, *-aū*, *-ī* presage the modern oppositions among masculine, neuter and feminine types such as Gujarati *navo*, *navū*, *navī* 'new', Hindi *nayā*, *naī* (m., f.). The case system of Apabhraṁśa is at a more advanced stage of disintegration than found earlier. For example, instrumental and locative plurals are now formally identical, and etymologically instrumental singular forms like *dāhiṇabhāē* are used in locatival function: *dāhiṇabhāē bharahu thakku* 'Bharata is located (*thakku*) in the southern division'. The paucity of distinct forms is evident in personal pronouns, where, for example, *maī*, *paī* (1st, 2nd person sg.) have functions equivalent to older accusative, instrumental and locative forms. Although Apabhraṁśa has some presents like *hoi* 'is', stems in *-a* of the type *kara-* 'do, make' (3 sg. *karai*) predominate. The Apabhraṁśa causative type *karāva-* (*karāvai*) is comparable to New Indo-Aryan formations (e.g. Gujarati *karāvɛ chɛ* 'has... do'). Moreover, Apabhraṁśa has causative formations found in modern languages but not attested earlier in Middle Indo-Aryan; e.g. *bhamāḍ-a-* 'cause to turn' (Gujarati *bhamāḍ-*).

The gender system of earlier Indo-Aryan is retained in some modern languages (e.g. Gujarati, see above), but is reduced in others (e.g. Hindi, with masculine and feminine only); some languages (e.g. Bengali) have eliminated systematic gender distinctions. Various inflectional forms are retained (e.g. Gujarati agentive *mɛ̃* 'I'), but the prevalent modern nominal system involves stems and postpositions or, much less commonly, prepositions. Over a large area of New Indo-Aryan, one finds variable nominals with direct and oblique forms, the former used independently, the latter with postpositions and other clitic elements. For example, Gujarati has singular direct forms in *-o* (m.), *-ū* (nt.), *-ī* (f.), oblique forms in *-ā* (m.-nt.), *-ī*. Some languages (e.g. Hindi) distinguish direct and oblique in the plural, others (e.g. Gujarati) do not. There are also nominals without these variations. Combinations of stems and postpositions serve the functions of inflected forms in earlier Indo-Aryan. Different languages have different postpositions for the same functions; e.g. Hindi *-ko*, Gujarati *-ne* mark definite direct objects, regularly animate, and indirect objects. Adjectives in general are formally like nouns, which they regularly precede in attributive

constructions, and, with few exceptions, postpositions follow such phrases, not individual components; e.g. Gujarati *mẽ tamārā dikrā-ne joyo* 'I saw your son'. Second person pronouns in New Indo-Aryan are differentiated essentially according to distinctions of deference, distance and familiarity, not according to number; e.g. Hindi *āp* has plural agreement but can refer to one person. Languages of the south-west also distinguish between first person inclusive and exclusive forms; e.g. Gujarati *ame* (exclusive), *āpṇe*. In demonstrative and relative pronouns, languages differ with regard to gender distinctions made; e.g. Marathi relative singular *jo* (m.), *je* (nt.), *ji* (f.), Gujarati *je* for all genders. They also differ in the deictic distinctions made.

The tendency to incorporate nominal forms in the verb system, evident in earlier times, continues into New Indo-Aryan. For example, Hindi has a contrast comparable to that of Bengali *korchi* 'am doing', *kori* 'do', both verbally inflected, but instead uses nominally inflected forms: *kar rahā/rahī hū* 'am doing', *kartā/kartī hū* 'do'. Gujarati lacks the contrast, but has verbally inflected presents (*karū chū* 'do, am doing') and nominally inflected preterits (*karto hato, kartī hatī*). Temporal auxiliaries like Hindi *hū*, Gujarati *chū* show verbal inflection, as do imperatives and some other forms. Person-number distinctions accord with the use of pronouns, but some languages (e.g. Bengali) have given up number distinctions in the verb. Future formations also show areal differences. Some languages have futures with *-š-* or *-h-* (e.g. Gujarati *kariš* 'I will do'), but *-b-* is characteristic of the east (e.g. Bengali *jabe* 'will go') and there are future formations that include gender distinctions, as in Hindi *jāegā* 'he will go', *jāegī* 'she will go'. The perfective of many New Indo-Aryan languages is semi-ergative, reflecting earlier participial constructions. For example, Gujarati *gher gayo/gaī* 'he/she went home' has masculine *gayo*, feminine *gaī*, depending on whether the agent is a man or a woman, but in *mẽ tamārā dīkrā-ne joyo* 'I saw your son' agreement (m. sg. *joyo*) is determined by the object (*dīkrā-ne* 'son'). Some languages (e.g. Hindi) suspend agreement if an object nominal takes a postposition, so that the construction is no longer strictly passive. A formal passive such as *nahī bulāyā jāegā* (m. sg.) 'will not be invited' in an example like Hindi *baccõ-ko nahī bulāyā jāegā* 'children will not be invited' is also construed with a noun phrase containing an object marker (*baccõ-ko*), so that this construction too is different from the passive of earlier Indo-Aryan. Moreover, formal passives normally are used in sentences without agent expressions except under particular semantic conditions; e.g. Gujarati *mārā-thī nahi jawāy* 'I (agentive *mārā-thī*) won't be able to go', with the passive *jaw-ā-y* (3 sg. pres.). As shown, formal passives are also not restricted to transitive verbs, and in some languages they are formed with a suffix, in others they are periphrastic formations.

Examples cited illustrate the usual unmarked word order of most New Indo-Aryan languages: subject (including agentive forms), object (with attributive adjectives, including number words, before this and preceded by possessives), verb (with auxiliaries). Adverbials can precede sentences or the verb. Relative clauses generally precede correlative clauses. A notable exception to the above, at least in its superficial order, is Kashmiri, where the verb occurs in second position.

Bibliography

Cardona and Emmerick (1974) contains a survey of Indo-Aryan on pages 439b–457a, including a table of languages and a map. Bloch (1965) is a general and masterful survey of the historical developments, while Varma (1972–6) is a handy summary of Grierson's survey of the modern languages, still valuable, though in serious need of updating. Turner (1966–9) is an indispensable reference work for lexicon, and includes an index by D.R. Turner.

References

Bloch, J. 1965. *Indo-Aryan from the Vedas to Modern Times* (Adrien-Maisonneuve, Paris; translation by A. Master, with revisions, additions and an index, of *L'Indo-Aryen du véda aux temps modernes*, Adrien-Maisonneuve, Paris, 1934)

Cardona, G. and R.E. Emmerick. 1974. 'Indo-Aryan Languages', in *The New Encyclopaedia Britannica: Macropaedia*, 15th ed., pp. 439b–457a (Encyclopaedia Britannica, Chicago)

Turner, R.L. 1966–9. *A Comparative Dictionary of the Indo-Aryan Languages*, 2 vols. (Oxford University Press, London)

Varma, S. 1972–6. *G.A. Grierson's Linguistic Survey of India, a Summary*, 3 vols. (Vishveshvaranand Institute, Panjab University, Hoshiarpur)

21 Sanskrit

George Cardona

1 Background

1.1 Introduction

Sanskrit (*saṁskṛta-* 'adorned, purified') refers to several varieties of Old Indo-Aryan, whose most archaic forms are found in Vedic texts: the *Rigveda* (*Ṛgveda*), *Sāmaveda*, *Atharvaveda*, *Yajurveda*, with various branches. Associated with these are groupings of explicatory and speculative works (called *brāhmaṇas*, *āraṇyakas*, *upaniṣads*) as well as texts concerning the performance of rites (*kalpa-* or *śrauta-sūtras*), treatises on phonetics, grammar proper, etymological explanations of particular words, metrics and astrology. Early Vedic texts are pre-Buddhistic — the composition of the *Rigveda* is plausibly dated in the mid-second millennium BC — although their exact chronology is difficult to establish. Brāhmaṇas and early sūtra works can properly be called late Vedic. Also of the late Vedic period is the grammarian Pāṇini (not later than early fourth century BC), author of the *Aṣṭādhyāyī*, who distinguishes between the language of sacred texts (*chandas*) and a more usual language of communication (*bhāṣā*, from *bhāṣ* 'speak'), tantamount to Classical Sanskrit. Epic Sanskrit is so called because it is represented principally in the two epics, *Mahābhārata* and *Rāmāyaṇa*. The date of composition for the core of early epic is considered to be in the first centuries BC. It is in the *Rāmāyaṇa* that the term *saṁskṛta-* is encountered probably for the first time with reference to the language. Classical Sanskrit is the language of major poetical works, dramas, tales and technical treatises on grammar, philosophy and ritual. It was not only used by Kalidasa and his predecessors but continued in use after Sanskrit had ceased to be a commonly used mother tongue. Sanskrit is a language of learned treatises and commentaries to this day. It has also undergone a literary revival, and original works are still being composed in this language. Indeed, Sanskrit is used as a lingua franca by paṇḍitas from different parts of India, and several thousand people claim it as their mother tongue.

1.2 Diachronic Changes Within Sanskrit

Linguistic changes are discernible in Sanskrit from earliest Vedic down to the language Pāṇini describes. The nominative plural masculine in *-āsas* (*devāsas* 'gods'), which has a counterpart in Iranian, is already less frequent in the *Rigveda* than the type in *-ās* (*devās*), and continues to lose ground; in Brāhmaṇas, *-ās* is the norm. The *Rigveda* has examples of an archaic genitive plural in *-ām* to *a*-stems, but the form in *-ānām* prevails here and is the only one used later. The instrumental singular of *a*-stems has both *-ā* and *-ena* (originally a pronominal type) in the *Rigveda* (*vīryā*/*vīryeṇa* 'heroic might, act'), but the latter is already prevalent and becomes the norm later. The Rigvedic nominative-accusative dual masculine of *a*-stems ends in *-ā* or *-au* (*mitrāvaruṇā*/*-varuṇau* 'Mitra and Varuṇa'), distributed according to phonological environments in early parts of the *Rigveda*, but *-au* steadily gains the upper hand and finally ousts *-ā* completely. For the nominative-accusative plural of neuter *a*-stems, the *Rigveda* has forms in *-ā* and *-āni: bhīmāni āyudhā* 'fearful weapons'. The former predominates in the *Rigveda*, but the situation is reversed in the *Atharvaveda*; later, *-āni* is the norm. Early Vedic had derivate *ī*-stems of two types, as in *vṛkīs* 'she wolf', *devī* 'goddess' (nom. sg.), *vṛkyas, devīs* (nom. pl.). The type *vṛkī-* is gradually eliminated as an independent formation, but leaves traces incorporated into the *devī* type (e.g. nom. pl. *devyas*). Rigvedic feminine *i*- and *u*-stems have instrumental singular forms of the type *ūtī* 'with, for help', *jātū* 'by nature' in addition to forms with *-ā* (*ūtyā*, *dhenvā* 'cow'). Even in the *Rigveda*, *u*-stems usually have forms of the type *dhenvā*, and the type *ūtyā* also becomes the norm later. Masculine and neuter stems in *-i*, *-u* have Rigvedic instrumental singulars with *-ā* (*pavyā*, *paśvā* to *pavi-* 'felly', *paśu-* 'animal') and *-nā* (*agninā* 'fire, Agni', *paśunā*). The latter predominate in the *Atharvaveda* and ultimately take over except for a few nouns (*patyā* 'husband', *sakhyā* 'friend'). The *Rigveda* has *avyas*, *madhvas*, genitive singulars of *avi-* 'sheep', *madhu-* 'honey'; the regular later forms are *aves*, *madhunas* (also *madhos* in Vedic). Endingless locatives like *ahan* (*ahan-* 'day') are also gradually eliminated in favour of forms with the ending *-i*: *ahani*/*ahni*. Early Vedic has pronominal forms not found in Classical Sanskrit: *asme*, *yuṣme* (loc. pl.) from the first and second person pronouns, replaced by *asmāsu*, *yuṣmāsu*; *āvos* (1st person gen.-loc. du.), *mahya* (1st person dat. sg.), replaced by *āvayos*, *mahyam*. Pāṇini expressly classes such earlier Vedic forms as belonging to the language of sacred texts.

The verbal system shows comparable differences. Early Vedic had modal forms from several stems: present, aorist, perfect. For example, the Rigvedic imperatives *śṛṇudhi*, *śṛṇuhi*, *śṛṇu* (2 sg.) and the Atharvavedic optative *śṛṇuyāt* (3 sg.) are formed to the present stem *śṛṇu-* of *śru* 'hear, listen', but the Rigvedic imperative *śrudhi* (2 sg.) and optative *śruyās* (3 sg.) are formed to the aorist stem. In later Sanskrit, imperatives and optatives regularly are formed from present stems. The first plural primary active

ending *-masi* (*bharāmasi* 'we carry'), which has an equivalent in Iranian, predominates over *-mas* in the *Rigveda*, but not in the *Atharvaveda*, and later *-mas* is the rule. Early Vedic forms like *ās* 'was' (3 sg. imperfect of *as*) and *avāṭ* (3 sg. aorist of *vah* 'transport') show the effects of the simplification of word-final clusters. Such forms are replaced by the types *āsīt*, *avākṣīt*, with *-īt* (2 sg. *-īs*), in which endings are clearly shown. Aorist forms made directly from verb roots are also replaced by forms from stems in *-a* or sigmatic stems, the latter especially in the medio-passive. Thus, the *Rigveda* has 1 sg. *akaram*, 2 sg. *akar* (< *akar-s*), 3 sg. *akar* (< *akar-t*), but the *Atharvaveda* has 2 sg. *akaras*, 3 sg. *akarat*, from *kṛ* 'make, do', and the *Rigveda* has not only a root aorist third plural middle *ayujran* but also a sigmatic form *ayukṣata* 'they yoked'. Commentators like Patañjali (mid-second century BC) and the etymologist Yāska before him used the sigmatic form *akṛṣata* (3 pl. middle) in paraphrasing a Vedic verse with the root aorist form *akrata*. Early Vedic forms of the type *śaye* 'is lying' are gradually replaced by the type *śete*, with *te*, which is explicitly marked for person.

Early Vedic distinguishes among the aorist, imperfect and perfect. The aorist is commonly used to refer to something that has recently taken place, and the imperfect is a narrative tense form used of acts accomplished or states prevailing at a past time not close at hand. For example, *úd u jyótir ... savitā́ aśret* 'Savitṛ has set up (*úd ... aśret*) the light (*jyótis*)', spoken at dawn, has the aorist *úd ... aśret*, but *ná mṛtyúr āsīd amṛ́tam ná tárhi ná rā́tryā áhna āsīt praketáḥ* 'then (*tárhi*) was there (*āsīt*) not (*ná*) death (*mṛtyús*) or deathlessness (*amṛ́tam*), nor was there the mark (*praketás*) of night (*rā́tryās*) or day (*áhnas*)' has the imperfect *āsīt*. The perfect originally signified, as in early Greek, a state of being; e.g. *bibhāya* '... is afraid'. From the earliest Vedic texts, however, this is not always the use of the perfect, which came to be used as a narrative tense. For example, the following Brāhmaṇa passage has both perfect and imperfects: *yajño vai devebhya ud akrāman na vo'ham annaṁ bhaviṣyāmīti/ neti devā abruvan annam eva no bhaviṣyasīti/ taṁ devā vimethire ... te hocur devā na vai na itthaṁ vihṛto'laṁ bhaviṣyati hantemaṁ yajñaṁ saṁ bharāmeti/ tatheti taṁ saṁ jabhruḥ* 'the sacrifice (*yajñas*) fled (*ud akrāmat*) from the gods (*devebhyas*), saying (citation particle *iti*), "I will not be (*na bhaviṣyāmi*) food (*annam*) for you (*vas*)"; the gods (*devās*) said (*abruvan*), "No, you will be (*bhaviṣyasi*) food for us (*nas*)"; the gods tore it apart (*taṁ vi methire*) ... the gods said (*ūcus*), "Truly (*vai*), it will not be sufficient (*na ... alaṁ bhaviṣyati*) for us thus (*ittham*) torn apart (*vihṛtas*), so let us put this sacrifice together (*imaṁ yajñaṁ saṁ bharāma*)"; they agreed (*tatheti* 'yes') and put it together (*taṁ saṁ jabhrus*)'. The imperfect *ud akrāmat, abruvan* and the perfect *vi methire, saṁ jabhrus* occur in similar contexts. This passage also illustrates the normal later combination of preverbs and verbs: preverbs immediately precede the verb stems with which they are connected; in earlier Vedic, tmesis was common — as in *úd ... aśret* of the Rigvedic passage cited earlier.

In addition, the augment became obligatory, as it had not been before, in imperfect and aorist forms.

The Brāhmaṇa passage just quoted also contains the future forms *bhaviṣyāmi, bhaviṣyasi, bhaviṣyati*, from the verb *bhū*, with the augmented suffix *-iṣya*. This and the unaugmented suffix *-sya* (*dāsya-* 'will give') are used from earliest Vedic on, but there is also a composite type, originally formed from an agent noun of the type *kartṛ́-* (nom.sg. *kartā́*) followed, except in the third person, by forms of the verb 'be': *kartāsmi* 'I will do', *kartāsi* 'you will do', *kartā* 'he will do'. This formation, which was in common use at Pāṇini's time, was rare in early Vedic. The perfect also has a periphrastic formation, for derived verbs such as causatives; e.g. *gamayāñ cakāra* (3 sg.) 'made to go' (3 sg. present *gamayati*), formed with the accusative singular of an action noun (*gamayā-*) and the perfect of *kṛ* 'do'. This type first appears in the *Atharvaveda* (form cited), and gains currency; Pāṇini recognises it not only as the regular perfect for derived verbs but also for some primitive verbs. Corresponding to future forms such as *bhariṣyati* 'will carry', there were, from earliest Vedic, secondary augmented forms like *abhariṣyat* 'was going to carry', and these are later to become the regular verbal constituents in contrary-to-fact conditional sentences.

Early Vedic has a category that goes out of use later: the injunctive, formally an unaugmented secondary form; for example, *bhūt, carat* are third person singular injunctives corresponding to the aorist *abhūt* and the imperfect *acarat*. In a *Rigveda* passage such as *agníh sáptiṁ vājambharáṁ dadāti ... agnír ródasī ví carat* 'Agni (*agnís*) gives (*dadāti*) a horse (*sáptim*) that carries away prizes (*vājambharám*) ... Agni wanders through (*ví carat*) the two worlds (*ródasī*)', the injunctive *ví carat* and the present *dadāti* are juxtaposed, both used of general truths. In such statements, Vedic also uses subjunctives, characterised by the vowel *-a-* affixed to a present, aorist or perfect stem, as in Rigvedic *ná duṣṭutí mártyò vindate vásu ná śrédhantaṁ rayír naśat* 'a mortal (*mártyas*) does not find (*ná vindate*) treasure (*vásu*) through bad praise (*duṣṭutí*), nor does wealth (*rayís*) come to (*naśat*) one who faulters in the performance of rites (*śrédhantam*)', where the present *vindate* is juxtaposed with the aorist subjunctive *naśat* 'reach'. In addition, subordinate clauses such as *pūṣā́ no yáthā ... ásad vṛdhé rakṣitā́* 'so that (*yáthā*) Pūṣan be (*ásat*) our protector in order that we might grow (*vṛdhé*)' use the subjunctive, which also occurs in requests; e.g. *devó devébhir ā́ gamat* 'may the god come (*ā́ gamat*) with the gods (*devébhis*)'. In negative commands, the injunctive is used with the particle *mā*, as in *mā́ no vadhīḥ ... mā́ párā dāḥ* 'do not kill (*mā́ vadhīs*) us (*nas*), do not forsake (*mā́ párā dās*) us', with the second person singular aorist injunctives *vadhīs, parā dās*. The regular negative particle used with a subjunctive, however, is *na*: e.g. *sá jáno ná reṣan máno yó asya ... ā vívāsāt* 'that person (*sá jánas*) does not suffer ill (*ná reṣat*), who seeks to win (*yás ā vívāsāt*) his (*asya*) spirit (*mánas*)' has the aorist subjunctive *reṣat* and the subjunctive of the present desiderative stem

ā vivāsa- (*-sāt* < *-sa-a-t*). Later, the injunctive is retained only in negative commands of the type *mā vadhīs*, 3 sg. *mā vadhīt*. The subjunctive also steadily loses ground until it is no longer current; for Pāṇini, subjunctive forms belong to the language of sacred texts. Only the first person type *karavāṇi* 'I may do, let me do', incorporated into the imperative system, is retained. The functions of the subjunctive are taken over by the optative and the future. For example, in Vedic a subordinate clause introduced by *yathā* may have a subjunctive or an optative, but *yadi* 'if' is regularly used with a subjunctive in early Vedic. Thus, a passage cited above has *yathā ... asat*, and *yáthā bhávema mīḷhúṣe ánāgāḥ* 'that we may be (*yáthā bhavema*) sinless (*ánāgās*) towards the gracious one (*mīḷhúṣe*)' has the optative *bhavema*, but *á gha gamad yádi śrávat* 'let him come (*á ... gamat*) if he hear (*yádi śrávat*)' has the aorist subjunctive *śravat*. In later Vedic, however, *yadi* is used with an optative, as in *yádi bibhīyā́d duścármā bhaviṣyāmīti somapauṣṇā́m śyāmā́m á labheta* 'if he fear (*yádi bibhīyā́t*) that he might be (*bhaviṣyāmīti* 'I will become') stricken by a skin disease (*duścármā* 'bad-skinned'), let him immolate (*á labheta*) a black goat (*śyāmā́m* 'black') dedicated to Soma and Pūṣan'.

Nominal forms within the verbal system of early Vedic are numerous. The Rigveda has derivatives with *-ya*, *-tva* that function as gerundives: *vācya-* 'to be said' (root *vac*), *kartva-* 'to be done' (*kṛ*). In addition, the *Atharvaveda* has forms with *-(i)tavya*, *-anīya*: *hiṁsitavya-* 'to be harmed', *upajīvanīya-* 'to be subsisted upon'. By late Vedic, the type with *-tva* has lost currency, and for Pāṇini the regular formations are of the types *kārya-*, *kartavya-*, *karaṇīya-*. In Indo-Aryan from Vedic down to modern times, gerunds are used with reference to the earlier of actions performed in succession, usually by the same agent ('after doing A, ... does B', '... does A before doing B'); e.g. *yuktvā́ háribhyāṁ úpa yāsad arvā́k* 'let him yoke his bay horses to his chariot (*yuktvā* 'after yoking') and come hither (*upa yāsad arvāk*) with them (*haribhyām* 'with two bay horses')', *gū́ḍhvī́ támo ... abodhi* '(dawn) has awakened (*abodhi*) after hiding away (*gū́ḍhvī*) the darkness (*támas*)', *piba niṣadya* 'sit down (*niṣadya* 'after sitting down') and drink (*piba*)'. The *Rigveda* has gerunds with *-tvā*, *-tvāya -tvī*, *-(t)ya*, but these are ultimately reduced to two main types: *-tvā* after simple verbs or verbs with the negative prefix *a(n)-*, *-ya* after compounds with preverbs. Early Vedic uses a variety of case forms of action nouns, including root nouns, as what western grammarians traditionally call infinitives; e.g. dat. sg. *vṛdhe* (root noun *vṛdh-* 'growing'), *-tave* (*dātave* 'to give'), gen. sg. *-tos* (*dātos*), the last two from a derivative in *-tu* which also supplies the accusative *-tum* (*dātum*). There are other Vedic types, but nouns in *-tu* are noteworthy in that for later Vedic the accusative with *-tum* and the genitive in *-tos*, the latter construed with *īś* or *śak* 'be able', become the norm. According to Pāṇini, forms in *-tum* and datives of action nouns are equivalent in sentences like *bhoktum/ bhojanāya gacchati* '...is going (*gacchati*) in order to eat'.

1.3 Sanskrit Dialects

That some formations fell into disuse in the course of Old Indo-Aryan is no surprise: the developments sketched above represent chronological and dialectal changes. Such changes were recognised by grammarians who spoke the language. Patañjali notes that second plural perfect forms like *cakra* or *ūṣa* (*vas* 'dwell) were not used in his time; instead, one used participial forms such as *kṛtavantas, ūṣitās* (nom. pl. m.). Grammarians also recognised that various dialects existed. Pāṇini takes note of forms used by northerners, easterners and various dialectal usages described by other grammarians. The etymologist Yāska notes, as does Patañjali, that finite forms of the verb *dā* 'cut' were used in the east, while in the north the verb occurred in the derivative *dātra-* 'sickle'. Earlier documents also afford evidence of dialect differences. The major dialect of the *Rigveda* is one in which Proto-Indo-European *l* merged with *r* (e.g., *pūrṇa-* 'full'), but other dialects developed *l*, and one finds doublets such as *rohita-/lohita-* 'red'. The development of retroflex liquids *-ḷ-*, *-ḷh-* from intervocalic *-ḍ- -ḍh-* is another characteristic of some areas, among them the major dialect of the *Rigveda*.

1.4 Sanskrit and Other Languages

Classical Sanskrit represents a development of one or more such Old Indo-Aryan dialects, accepted as standard, at a stage when archaisms such as those noted (section 1.2) had largely been eliminated. It is plausible to accept that both Classical Sanskrit and earlier dialects of Indo-Aryan coexisted with vernaculars that were removed from these by changes which characterise Middle Indo-Aryan, just as in later times Sanskrit and vernaculars were used side by side under particular circumstances. There is evidence to support this view, particularly in Patañjali's *Mahābhāṣya*, where he discusses the use of 'correct speech forms' (*śabda*) and 'incorrect speech forms' (*apaśabda*), considered corruptions (*apabhraṁśa*) of the former. Patañjali speaks of *śiṣṭas*, model speakers, who are characterised as much by moral qualities as by their speech. They are Brāhmaṇas who reside in Āryāvartta, the land of the Āryas in north-central India, who at any time have only as much grain as will fit in a small pot, who are not greedy, who behave morally without ulterior motives and who attain full knowledge of traditional learning with consummate ease, not having to be taught. These model speakers are those one should imitate and, it is assumed, the models Pāṇini followed in composing his grammatical rules. However, even learned men did not avoid vernaculars, as Patañjali also points out. He remarks that a restriction such that correct speech forms should be used to the exclusion of others is absolute only in respect of rituals. To illustrate, Patañjali speaks of sages who said *yar vā ṇaḥ* 'what is ours', *tar vā ṇaḥ* 'that is ours' instead of *yad vā naḥ, tad vā naḥ* but did not use such forms in the course of ritual acts. Now, forms like *yar* instead of *yad* reflect an Indo-Aryan tendency to eliminate obstruence for non-initial retroflex and dental stops; the particular

change in question is seen also in Prakrit *bāraha* as opposed to Sanskrit *dvādaśa* 'twelve'. Moreover, Patañjali must have been, if not a native speaker of Sanskrit in the strictest sense, at least one fully fluent in the language, with authority concerning its usage. For he explicitly distinguishes between what is desirable — that is, what is required by accepted usage — and what obtains by grammatical rules. At Patañjali's time, then, Sanskrit must have been a current vehicle of communication in certain circles and under particular social and religious conditions, used concurrently with vernaculars. Much the same picture is painted for later periods, when Sanskrit was doubtless revived. Thus, in his *Kāmasūtra*, Vātsyāyana notes that to be held in high esteem a man-about-town should use neither Sanskrit nor a local language exclusively. Indeed, the coexistence of Middle Indo-Aryan and Sanskrit speech is to be envisaged even for the time when very early texts were given their final redactions. The *Rigveda* has forms like *vikaṭa-* 'deformed' and *jyotis-* 'light'. The former is a Middle Indo-Aryan form of *vikr̥ta-*, with *-aṭ-* for *-r̥t-*, comparable to Aśokan *kaṭa-* 'made' (Skt. *kr̥ta*), and the latter had *jy-* for *dy-*. It has been suggested, plausibly in my estimation, that there was an archaic Middle Indo-Aryan contemporaneous with early Vedic.

Sanskrit was also subject to non-Aryan influence from early on. In the sixth century BC Darius counted Gandhara as a province of his kingdom, and Alexander the Great penetrated into the north of the subcontinent in the fourth century. From Iranian come terms such as *lipi-* 'writing, script', *kṣatrapa-* 'satrap', and Greek is the source of such words as *kendra-* 'centre', *jāmitra-* 'diameter', *horā-* 'hour'. At a later time borrowings entered from Arabic and other sources. But long before this Sanskrit was influenced by Dravidian, from which it borrowed terms such as *kāla-* 'black', *kuṭ̆i-* 'hut' (cf. Tamil *kaṛ* 'blackness', *kuṭi*) and the influence of which contributed to the spread of retroflex consonants (see section 1 of the chapter on Indo-Aryan). It is not certain in every instance, however, that borrowing proceeded from Dravidian to Indo-Aryan, since Dravidian languages also freely borrowed from Indo-Aryan. For example, some scholars maintain that Skt. *kaṭu-* 'sharp, pungent' is a Dravidian borrowing, but others treat it as a Middle Indo-Aryan development of **kr̥tu-* 'cutting' (root **kr̥t* 'cut'). Whatever be the judgement on any individual word, nevertheless, it is clear that Sanskrit and other Indo-Aryan dialects borrowed from Dravidian sources.

2 Brief Description of Classical Sanskrit

2.1 Sound System and Script

The sounds of Sanskrit are shown in table 21.1. In the present context, it is not necessary to take a particular stand about which sounds should be considered 'basic', 'underlying' or 'phonemic'. Suffice it to note that sounds

Table 21.1 The Sounds of Sanskrit

Vowels

		i	ī			u	ū	
			e				o	
				a				
				ā				
	r̥	[r̥̄]	[l̥]			ai	au	

Consonants

	Obstruents Voiceless		*Obstruents* Voiced		*Nasals*	*Semi-vowels*	*Liquid*	*Tap*	*Spirants* Voiceless	*Spirants* Voiced
Pharyngeal									[ḥ]	h
Velar	k	kh	g	gh	[ṅ]				[χ]	
Palatal	c	ch	j	jh	[ñ]	y			ś	
Retroflex	ṭ	ṭh	ḍ	ḍh	ṇ			r*	ṣ	
Alveolar								r*		
Dental	t	th	d	dh	n		l		s	
Labio-dental						v				
Labial	p	ph	b	bh	m [ṁ]				[φ]	

Note: *Some ancient authorities say *r* is retroflex, others say it is alveolar.

of table 21.1 within square brackets have restricted distributions. r̥̄ occurs only in accusative or genitive plurals of r̥-stems (*pitr̥̄n* 'fathers', *mātr̥̄s* 'mothers', gen. pl. *pitr̥̄ṇām*, *mātr̥̄ṇām*, rare nom.-acc. pl. nt. *kartr̥̄ṇi* 'which do'); l̥ is found only in forms of *kl̥p* 'be fit, arrange, imagine' (past participle *kl̥pta-*). Due to the reduction of word-final clusters, *-ṇ* occurs in words such as *prāñ* (nom. sg.) 'directed forward, toward the east', but otherwise *ṇ* and *ñ* are found before velar and palatal stops, respectively, though not necessarily as replacements of *n* or *m* at morph boundaries. The nasal off-glide *ṁ* occurs word-internally before spirants at morph boundaries as the final segment of items that have *-n* or *-m* before vowels and in word-final position before spirants and semi-vowels or stops, where it varies with nasalised semi-vowels and nasal stops homorganic with following stops. *ḥ* is a word-final segment in prepause position or before voiceless spirants, velars and labials. *χ* φ are alternants to *-ḥ* before velars and labials. Like *ṅ* and *ñ*, *ṇ* is not the initial sound of lexical items. It occurs in word-final position, though rarely except before nasals as the final sound of a morph that has a non-nasal retroflex stop before vowels, but intervocalic *-ṇ-* is found in words like *kaṇa-* 'grain, atom', that do not contain sounds which condition retroflexion.

The vowels *i*, *u* and *ī*, *ū* differ essentially in duration: short vowels last one mora (*mātrā*), long vowels two morae; however, in accepted modern pronunciations, *i* and *u* can be lower than their long counterparts. *e*, *o* are monophthongs of two morae, though they derive historically from diphthongs and alternate with *ay*, *av* before vowels. *ai*, *au* are diphthongs for

which ancient phoneticians and grammarians recognised dialect variants: for example, the first segment of each was a closer vowel in some dialects than in others. Prosodically, however, *ai*, *au* behave in the manner of simple long vowels, and there are good reasons for not treating them as combinations of *ā* with *i*, *u*. *ṛ* is also a complex sound, consisting of *r* surrounded by vowel segments, according to a fairly old description, but this also behaves prosodically as a single vowel. In north-central India, *ṛ* is pronounced as *r* followed by short *i*. *a*, *ā* behave as a pair of short and long vowels, but they are also qualitatively different, as shown. Vowels can be unnasalised or nasalised. They also have pitch differences such that they are called *anudātta*, *udātta* and *svarita*. Pāṇini's statements concerning these are best understood as reflecting a system in which an anudātta vowel is low-pitched, an udātta vowel is high-pitched, and a svarita vowel has a combination of both pitches: *a*, *á*, *à*. According to Pāṇini, a svarita vowel is high-pitched for the duration of half a mora from its beginning, low-pitched for its remainder, but there were dialectical variations, as can be seen from other ancient descriptions. There are also differences in Vedic traditions of recitation concerning the relative pitches of the vowels in question.

Sanskrit generally does not allow word-final clusters, although *-r*C is permitted if both consonants belong to the same element; e.g. *ūrk* (nom. sg.) 'strength' (acc. sg. *ūrj-am*). Sanskrit also has a fairly complex system of morphophonemic adjustments (*sandhi*) across grammatical boundaries, at word boundaries if the items in question are pronounced in close juncture (*saṁhitāyām*). Some of these adjustments are illustrated in examples given; e.g. in the Brāhmaṇa passage cited in section 1.2: *yajño vai* ← *yajnas vai*, *devebhya ud* ← *devebhyas ud*, *akrāman na* ← *akrāmat na*, *voham* ← *vas aham*, *annaṁ bhaviṣyāmīti* ← *annam bhaviṣyāmi iti*, *neti* ← *na iti*, *devā abruvan* ← *devās abruvan*, *no bhaviṣyasi* ← *nas bhaviṣyasi*, *taṁ devā vi methire* ← *tam devās vi methire*, *hocur devā na* ← *ha ūcus devās na*, *tatheti* ← *tathā iti*, *taṁ saṁ jabhruḥ* ← *tam sam jabhrus*, the last with *-ḥ* instead of *-s in pausa*. These adjustments also affect vowel pitches. The particular place of a high-pitched vowel in an underived base is not predictable. In general, a syntactic word has one high-pitched vowel only — but may have none — and a finite verb form following a term that is not a finite verb has no high-pitched vowel except in particular collocations. Further, a low-pitched vowel following a high-pitched one shifts to a svarita vowel, as in *á gàmat* ← *á gamat*. There are other accentual adjustments that involve considerable complexity and dialectal variation.

Sanskrit was and continues to be written in various scripts in different areas, but the most widely recognised is the Devanāgarī script, the symbols of which are shown in table 21.2. These are traditionally arranged as follows: symbols for vowels, then for consonants; the latter are subdivided into: stops (five groups of five), semi-vowels, voiceless spirants, *h*. In addition, there are symbols for *ḷ* and *ḥ*. *ṁ* is designated by a dot (*bindu*) over a consonant or a

Table 21.2 Devanāgarī Symbols and their Transliterations

Vowels (*svarāḥ*)

अ	आ	इ	ई	उ	ऊ	ऋ	ॠ	ऌ	ए	ऐ	ओ	औ
a	ā	i	ī	u	ū	ṛ	ṝ	ḷ	e	ai	o	au

Consonants (*vyañjanāni*)

Stops (*sparśāḥ*)					*Semi-vowels* (*antaḥsthāḥ*)	*Spirants* (*ūṣmāṇaḥ*)		*Others*	
क	ख	ग	घ	ङ				ह	ः
k	kh	g	gh	ṅ				h	ḥ
च	छ	ज	झ	ञ	य	श			
c	ch	j	jh	ñ	y	ś			
ट	ठ	ड	ढ	ण	र	ष			ळ
ṭ	ṭh	ḍ	ḍh	ṇ	r	ṣ			ḷ
त	थ	द	ध	न	ल	स			
t	th	d	dh	n	l	s			
प	फ	ब	भ	म	व				
p	ph	b	bh	m	v				

Examples of combinations

का	काँ	कि	की	कु	कू	कृ	कॄ	कॢ	क्त	क्र	क्ष	ज्ञ	त्र	त्व	द्य
kā	kāṁ	ki	kī	ku	kū	kṛ	kṝ	kḷ	kta	kra	kṣa	jña	tra	tva	dya

द्र	द्व	प्त	ब्द	र्क	र्कं	श्च	श्र	श्व	स्त	स्य	स्र	स्व	ह्म
dra	dva	pta	bda	rka	rkaṁ	śca	śra	śva	sta	sya	sra	sva	hma

ह्य	ह्र	ह्ल	ह्व	त्स्न्य
hya	hra	hla	hva	rtsnya

Numerals

१	२	३	४	५	६	७	८	९	०
1	2	3	4	5	6	7	8	9	0

Note: I have adopted the most generally accepted order of symbols and the subgroupings most widely accepted traditionally; the usual Sanskrit terms for sound classes are given in parentheses.

vowel symbol, nasalisation by a dot within a half-moon (*ardhacandra*) over a symbol; *χ φ* are designated by *ˣ* before symbols for voiceless velars and labials.

In referring to vowels, one pronounces the sounds in question; e.g. '*a*' denotes the vowel *a*. Consonants in general are referred to by a combination of the sounds and a following *a*: e.g., '*ka*' denotes *k*. In addition, a sound name is formed with suffixed *-kāra*; e.g., '*akāra*', '*kakāra*' refer to *a*, *k*.

Certain sounds, however, have particular names: *r ḥ ṁ χ φ*, respectively, are called *repha*, *visarjanīya* (or *visarga*), *anusvāra*, *jihvāmūlīya*, *upadhmānīya*.

Consonant symbols, except those for *ḥ ṁ χ φ*, without any appended element, denote consonants followed by *a*. Other consonant-vowel combinations are designated by consonant symbols with appended vowel symbols, which may precede, follow, or come under the former, as illustrated in table 21.2. There are also ligatures for consonant combinations, some of which are illustrated in table 21.2. Finally, there is a set of Devanāgarī numerals. Variants of symbols are found in different areas.

2.2 Grammar

2.2.1 Introduction

Although many archaic features of earlier Vedic dialects have been eliminated in Sanskrit, the grammatical system nevertheless remains quite rich. Singular, dual and plural forms are distinguished in both the nominal and the verbal systems, and ablaut variations are maintained in many types of formations.

2.2.2 Nominal system

Eight cases can be distinguished, although the vocative does not have a syntactic status comparable to the others: nominative (nom.), vocative (voc.), accusative (acc.), instrumental (inst.), dative (dat.), ablative (abl.), genitive (gen.), locative (loc.), according to traditional western terminology. All eight are formally distinguished in the singular of masculine *a*-stems; e.g. *deva-* 'god': nom. *devas*, voc. *deva*, acc. *devam*, inst. *devena*, dat. *devāya*, abl. *devāt*, gen. *devasya*, loc. *deve*. Otherwise, there are homophonous forms as follows. All stems: dual nom.-voc.-acc., inst.-dat.-abl., gen.-loc.: *deva-*: *devau*, *devābhyām*, *devayos*; *phala-* (nt.) 'fruit': *phale*, *phalābhyām*, *phalayos*; *senā-* (f.) 'army': *sene*, *senābhyām*, *senayos*; *agni-* (m.) 'fire': *agnī*, *agnibhyām*, *agnyos* (similarly *kṛti-* (f.) 'deed'); *vāri-* (nt.) 'water': *vāriṇī*, *vāribhyām*, *vāriṇos*; *vāyu-* (m.) 'wind': *vāyū*, *vāyubhyām*, *vāyvos* (similarly *dhenu-* (f.) 'cow'); *madhu-* (nt.) 'honey': *madhunī*, *madhubhyām*, *madhvos*; *devī-* 'goddess': *devyau*, *devībhyām*, *devyos*; *vadhū-* 'bride': *vadhvau*, *vadhūbhyām*, *vadhvos*; *sakhi-* (m.) 'friend': *sakhāyau*, *sakhibhyām*, *sakhyos*; *pitṛ-* 'father': *pitarau*, *pitṛbhyām*, *pitros* (similarly *mātṛ-* 'mother'); *kartṛ-* 'doer, maker': *kartārau* (m.) *kartṛṇī* (nt.), *kartṛbhyām*, *kartros*; *go-* 'ox, cow': *gāvau*, *gobhyām*, *gavos*; *rājan-* 'king': *rājānau*, *rājabhyām*, *rājños*; *vāc-* (f.) 'voice, speech': *vācau*, *vāgbhyām*, *vācos*; *sraj-* (f.) 'garland': *srajau*, *sragbhyām*, *srajos*; nom.-voc. pl.: *devās*, *phalāni*, *senās*, *agnayas*, *kṛtayas*, *vārīṇi*, *vāyavas*, *dhenavas*, *madhūni*, *devyas*, *vadhvas*, *sakhāyas*, *pitaras*, *mātaras*, *kartāras kartṝṇi*, *gāvas*, *rājānas*, *vācas*, *srajas*. All stems except personal pronouns:

dat.-abl. pl.: *devebhyas*, *phalebhyas*, *senābhyas* etc. (with *agni-* etc. and *-bhyas*), *rājabhyas*, *vāgbhyas*, *sragbhyas*, but dat. *asmabhyam* 'us', *yuṣmabhyam* 'you', abl. *asmat*, *yuṣmat*. Nom.-acc. of all numbers for neuter stems: sg. *phalam*, *vāri*, *madhu*, *kartṛ*; for dual and plural see above. Abl.-gen. sg. except for masculine and neuter *a*-stems and personal pronouns: *senāyās*, *agnes*, *kṛtes/kṛtyās*, *vāriṇas*, *dhenos/dhenvās*, *madhunas*, *devyās*, *vadhvās*, *sakhyus*, *pitus*, *mātus*, *kartus*, *gos*, *rājñas*, *vācas*, *srajas*, but *devāt devasya* (similarly for *phala-*), *mat mama*, *tvat tava*. The accusative plural of feminine *ā*-stems and consonant stems is homophonous with the nominative and vocative plural (see above), but other stems make a distinction: *devān*, *agnīn*, *kṛtīs*, *vāyūn*, *dhenūs*, *devīs*, *vadhūs*, *sakhīn*, *pitṝn*, *mātṝs*, *kartṝn*, *rājñas*. In the singular, a few stems make no distinction between nominative and vocative (e.g. *gaus*, *vāk*, *śrīs* 'splendour, wealth'), but the two are usually distinguished: *devas*, *deva*; *senā*, *sene*; *agnis*, *agne*; *kṛtis*, *kṛte*; *vāri*, *vāre/vāri*; *vāyus*, *vāyo*; *dhenus*, *dheno*; *madhu*, *madho/madhu*; *devī*, *devi*; *vadhūs*, *vadhu*; *sakhā*, *sakhe*; *pitā*, *pitar* (similarly *mātṛ-*, *kartṛ-*); *rājā*, *rājan*. As can be seen, certain endings have variants according to stems, and this is true of the genitive plural, which has *-ām* after consonant stems (*rājñām*, *vācām*, *srajām*) and some vowel stems (e.g. *śriyām*, *gavām*) but *-nām* after most vowel stems, with lengthening of short vowels before this ending: *devānām*, *phalānām*, *senānām*, *agnīnam* etc.; however, personal pronouns have *-kam* (*asmākam*, *yuṣmākam*), and other pronominals have *-sām* (e.g. *teṣām* 'of them').

Endings are divisible into two groups with respect to phonological and grammatical alternations; nominative, vocative, accusative singular and dual and nominative plural for non-neuter stems as well as the nominative and accusative plural for neuter stems are strong endings, others are weak endings. Consonant-initial weak endings behave phonologically as though they were separated from stems by a word boundary; for example, *as*-stems have variants with *-o* before *-bhyām* (inst.-dat.-abl. du.), *-bhis* (inst. pl.), *-bhyas* (dat.-abl. pl.), *-aḥ* before *-su* (loc. pl.): *manas-* 'mind, spirit': nom.-acc. sg. *manas*, inst. sg. *manasā* but *manobhyām*, *manobhis*, *manaḥsu*.

Stems show variation that in part reflects Proto-Indo-European ablaut alternation. For example: *agni/agne-* (*agnay-* before vowels), *vāyu-/vāyo-* (*vāyav-*), *sakhi-/sakhe-/sakhāy/sakhā-*, *pitṛ-/pitar-/pitā-*, *kartṛ-/kartar-/kartār-/kartā-*, *rājan-/rājān-/rājā-/rājñ-* (before vocalic weak endings)/*rāja-* (before consonantal weak endings). There are also heteroclitic stems such as *asthi-/asthan-* (nt.) 'bone': nom.-acc. sg. *asthi*, du. *asthinī*, pl. *asthīni*, inst.-dat.-abl. du. *asthibhyām*, etc., with *asthi-* before consonantal weak endings, but inst. sg. *asthnā* etc., with *asthn-* before vocalic weak endings, and loc. sg. *asthani/asthni*. Due to the palatalisation of *k*, *g* to *c*, *j* before front vowels prior to the merger of *ĕ* with *ă* and to analogic realignments, there are stems with palatals before vocalic endings and velars elsewhere; e.g. *vāc-*, *sraj-* (see above).

Adjectives generally pattern in the manner of comparable nouns. For example, *śukla-*, *śuklā-* 'white', *śuci-* 'bright', *guru-* 'weighty, heavy', *paṅgū-* 'lame' inflect in the same way as noun stems in *-a*, *-ā*, *-i*, *-u*, *-ū*. There are also consonant stem adjectives with ablaut alternation; e.g. *sant-/sat-* 'being' (m. nom. sg. *san*, nom.-acc. du. *santau*, nom. pl. *santas*, acc. sg. *santam*, acc. pl. *satas*, inst. sg. *satā*, inst.-dat.-abl. du. *sadbhyām*, etc.), *gacchant-/gacchat-* 'going' (*gacchan*, *gacchantau*, *gacchantas*, *gacchantam*, *gacchatas*, *gacchatā*, *gacchadbhyām*, etc.), *vidvans-/vidvāns-/viduṣ-/vidvad-* 'one who knows' (*vidvān*, *vidvan* (voc. sg.), *vidvāṁsau*, *vidvāṁsas*, *vidvāṁsam*, *viduṣā*, *vidvadbhyām*, etc.). In addition, there are adjectives that inflect pronominally. For example, nom. pl. *sarve*, dat. sg. *sarvasmai* (m.-nt.), *sarvasyai* (f.), gen. pl. *sarveṣām*, *sarvāsām*, from *sarvă-* 'whole, all', are comparable to *te*, *tasmai*, *tasyai*, *teṣām*, *tāsām* from *tă-* 'this, that'.

Personal pronouns not only have variants but also distinguish between independently accented and enclitic forms: acc. sg. *mā tvā*, dat. sg. *me te*, acc.-dat.-gen. du. *nau vām*, acc.-dat.-gen. pl. *nas vas* are enclitics corresponding to sg. acc. *mām tvām*, dat. *mahyam tubhyam*, gen. *mama tava*, du. acc. *āvām yuvām*, dat. *āvābhyām yuvābhyām*, gen. *āvayos yuvayos*, pl. acc. *asmān yuṣmān*, dat. *asmabhyam yuṣmabhyam*, gen. *asmākam yuṣmākam*. Demonstrative pronouns distinguish various degrees of proximity and distance: *etad* 'this here', *idam* 'this', *tad* 'this, that', *adas* 'that yonder' (all nom.-acc. sg. nt.). Interrogative and relative pronouns respectively have *kă-*, *yă-*, which inflect like pronominal *a*-stems except in the nominative and accusative singular neuter of the former (*kim yad*).

The Sanskrit system of number words is a familiar Indo-European one in that terms for 'one' to 'four' show inflectional and gender variation, but it also differs from the system of other ancient Indo-European languages in that higher number words also inflect; e.g. inst. pl. *pañcabhis* 'five', *ṣaḍbhis* 'six', *saptabhis* 'seven', *aṣṭābhis* 'eight', *navabhis* 'nine', *daśabhis* 'ten'.

Sanskrit is also like other older Indo-European languages in using suffixes for deriving what are traditionally called comparatives and superlatives, with two kinds of suffixes. For example, *garīyas-* 'quite heavy', *gariṣṭha-* 'exceedingly heavy' have *-īyas* and *-iṣṭha* following *gar-*, a form of the base that appears in the adjectival derivative *guru-*, but *-tara* and *-tama* follow adjectival stems, as in *madhumattara-* 'quite sweet', *madhumattama-* 'exceedingly sweet', from the stem *madhumat-*. It is noteworthy that *-tara*, *-tama* are used not only in derivates like *uttara-* 'upper, superior', *uttama-* 'highest', from *ud* 'up', but also in derivates from terms like *na* 'not' and finite verb forms: *natarām* 'the more not so (in view of an additional argument)', *natamām* 'all the more not so', *pacatitarām* 'cooks quite well', *pacatitamām* 'cooks exceedingly well'.

Derived nominal bases formed directly from verb roots include action nouns like *gati-* 'going', *pāka-* 'cooking', agent nouns such as *kartr̥-*, *kāraka-* 'doer, maker', object nouns like *karman-* 'deed, object', instrument nouns

such as *karaṇa-* 'means', participles like *gata-* 'gone', *kr̥ta-* 'done, made', gerunds, gerundives and abstract nouns that function as infinitives (see section 1.2). Bases with secondary derivate affixes (*taddhita* affixes) are of several types. There is a large group of derivates that correspond to phrases of the type *X-E Y-*, with which they alternate, where the values of *X-E* are case forms of particular nominals and *Y* stands for a nominal whose meaning is attributable to the derivational affix. For example, there are patronymics such as *dākṣi-* 'son of Dakṣa': any case form of *dākṣi-* corresponds to and alternates with a phrase containing the genitive *dakṣasya* 'of Daksa' and a form of *putra-* 'son' or a synonym. Other derivatives are formed from a more restricted set of nominals — predominantly pronominals — and correspond to particular case forms; e.g. *tatas* 'from that, thence', *tatra* 'in that, there' correspond respectively to ablative and locative forms of *tad-* 'this, that', with which they alternate. There are also redundant affixes. For example, *aśvaka-* 'nag' differs in meaning from *aśva-* 'horse', but *avika-* and *avi-* 'sheep' show no such semantic difference. Moreover, some taddhita affixes form derivates which do not alternate with forms or phrases containing items to which they are added. Thus, *kr̥trima-* 'artificial' has a suffix *-ma*, but *kr̥trima-* does not alternate with a phrase containing a form of *kr̥tri-*, since there is no such action noun: once *-tri* is affixed to *kr̥*, then, *-ma* is obligatory.

Compounds are of four general types: tatpuruṣa (determinative), dvandva (copulative), bahuvrīhi (exocentric), and a type that is usually invariant (avyayībhava). The first member of a tatpuruṣa compound is generally equivalent to a case form other than a nominative. For example, *tatpuruṣas* (nom. sg. m.) 'his man, servant' is equivalent to *tasya puruṣas*, with which it can alternate. Similarly, *grāmagatas* 'gone to the village' is equivalent to *grāmaṁ gatas*, with the accusative *grāmam* 'village'. There is a subtype of tatpuruṣa compounds in which the first member is coreferential with the second, which it modifies, as in *nīlotpalam* 'blue (*nīla-*) lotus', equivalent to *nīlam utpalam*, with two nominatives. Copulative compounds are equivalent to phrases with *ca* 'and'; e.g. *mātāpitarau* 'mother and father' alternates with *mātā pitā ca*. The term *bahuvrīhi* is an example of a bahuvrīhi compound: *bahuvrīhis* is equivalent to *bahur vrīhir asya*, used with reference to someone who has (*asya* 'of this') much (*bahus*) rice (*vrīhis*); similarly: *prāptodaka-* '(somewhere) that water (*udaka-*) has reached (*prāpta-*)', *ūḍharatha-* '(an animal) by which a chariot (*ratha-*) has been drawn (*ūḍha-*)'. There are also exocentric compounds which, for technical reasons, belong to the tatpuruṣa group; e.g. *pañcagava-* 'a group of five cows', a member of the subgroup of tatpuruṣas called *dvigu*. Avyayībhava compounds are generally, though not always, invariant; e.g. *upāgni* 'near the fire', *anujyeṣṭham* 'according to (*anu*) seniority (*jyeṣṭha-* 'oldest')'. Compounds like *upāgni* do not have alternative phrases containing the members of the derivate.

2.2.3 Verbal System

The basic elements on which the Sanskrit verbal system is built are the verb base or root, either primary or derived, and the present-imperfect stem. The root is the base for the present-imperfect stem, for various aorist stems and future formations, the perfect, the conditional and the precative. The present-imperfect stem is the basis not only for present and imperfect forms but also for imperative and optative forms. Although Sanskrit has eliminated quite a few complexities found in Vedic, its verbal system is still varied.

There is a systematic contrast between active and medio-passive. Some verbs take only active endings in agentive forms, others only middle endings. For example, the present *asmi*, *asi*, *asti* (1, 2, 3 sg.), *svas*, *sthas*, *stas* (1, 2, 3 du.), *smas*, *stha*, *santi* (1, 2, 3 pl.) and the imperfect *āsam āsīs āsīt*, *āsva āstam āstām*, *āsma āsta āsan* have only active endings with *as* 'be', and *āse āsse āste*, *āsvahe āsāthe āsāte*, *āsmahe ādhve āsate*, *āsi āsthās āsta*, *āsvahi āsāthām āsātām*, *āsmahi ādhvam āsata* have middle endings with *ās* 'be seated'. Other verbs take either active or middle endings in agentive forms, depending on a semantic contrast: if the result of the act in question is intended for the agent, middle endings are used, if not, active endings occur. For example, *kurute* is used with reference to someone making something for himself, *karoti* of one making something for another. Medio-passive endings alone are used in passives; e.g. *kaṭaḥ kriyate* 'a mat (*kaṭas*) is being made', with *-te* after the passive stem *kriya-*. Sanskrit also has formally passive forms comparable to the impersonal middle found in other Indo-European languages (the type Latin *itur* 'it is gone' i.e. 'one goes'), but it allows an agent to be signified with an instrumental in construction with such forms; e.g. *devadattena supyate* 'Devadatta is sleeping', with the formally passive *supyate* (act. *svapiti*) and the agentive instrumental *devadattena*. In both active and middle sets, three groups of endings are distinguished, which, following usual western terminology, I shall call primary, secondary and perfect endings. Although comparative evidence shows that certain primary endings were originally complexes with a particle, analogic developments have obscured this relation in some instances. The contrast between primary and secondary endings has been illustrated above: primary active; *-mi*, *-si* (*asi* < *as-si*), *-ti*; *-vas*, *-thas*, *-tas*; *-mas*, *-tha*, *-anti/ati* (e.g. *juhvati* 'they offer oblations'); secondary active: *-am*, *-s*, *-t* (augmented *-īs -īt*); *-va*, *-tam*, *tām*; *-ma*, *-ta*, *-ant/us* (e.g. *ajuhavus* 'they offered oblations', *adus* 'they have given', *akārṣus* 'they have made'); primary medio-passive: *-e*, *-se*, *-te*; *-vahe*, *-āthe*, *-āte*; *-mahe*, *-dhve* (*ādhve* < *ās-dhve*), *-ate/ante* (e.g. *edhante* 'they thrive'); secondary medio-passive: *-i*, *-thās*, *-ta*; *-vahi*, *-āthām*, *-ātām*; *-mahi*, *-dhvam*, *-ata/anta*. Certain endings are particular to the perfect, as can be seen from the following (*kṛ*): active: *cakăr-a*, *cakar-tha*, *cakār-a*; *cakṛ-va*, *cakr-athus*, *cakr-atus*; *cakṛ-ma*, *cakr-a*, *cakṛ-us*; medio-passive: *cakr-e*, *cakṛ-ṣe*, *cakr-e*; *cakṛ-vahe*, *cakr-āthe*, *cakr-āte*; *cakṛ-mahe*, *cakṛḍhve*, *cakr-ire*.

There is also a contrast between augmented and unaugmented stems. Indicative imperfect and aorist forms, as well as those of the conditional, have augmented stems. The augment is *a* for consonant-initial bases, *ā* for vowel-initial bases; e.g. imperfect *akarot*, aorist *akārṣīt*, conditional *akariṣyat* from *kṛ*, imperfect *āsit* (3 pl. *āsan*) from *as*.

Present-imperfect stems may be considered according to two major criteria. Some stems consist simply of verb roots, others have affixes; some stems exhibit grammatical alternation (ablaut), others do not. Stems that do not show grammatical alternation regularly have suffixes with *-a*: root-accented *bhav-a-* 'be, become' (*bhavāmi*, *bhavasi*, *bhavati*; *bhavāvas*, *bhavathas*, *bhavatas*; *bhavāmas*, *bhavatha*, *bhavanti*); *edh-a-* 'thrive' (*edhe*, *edhase*, *edhate*; *edhāvahe*, *edhethe*, *edhete*; *edhāmahe*, *edhadve*, *edhante*); *dīv-ya-* 'gamble' (*dīvyāmi* etc.); suffix-accented *tud-a-* 'goad, wound' (*tudāmi* etc.), passive *kri-ya-*. Such stems have *-ā* (< **o* by 'Brugmann's Law') before *-v-*, *-m-* of endings and *-e-* in second and third dual medio-passive forms. Root presents generally exhibit ablaut variation: full-grade in the singular active indicative, zero-grade elsewhere. For example: *as-ti*, *s-tas*, *s-anti*; *han-ti*, *ha-tas*, *ghn-anti* (*han* 'kill'); *dveṣ-ṭi*, *dviṣ-ṭas*, *dviṣ-anti*; *dviṣ-ṭe*, *dviṣ-āte*, *dviṣ-ate* (*dviṣ* 'hate'); *dog-dhi*, *dug-dhas*, *duh-anti*; *dug-dhe*, *duh-āte*, *duh-ate* (*duh* 'milk'). On the other hand, *ad* 'eat' has an invariant root stem (*at-ti at-tas ad-anti*) due in the first instance to phonologic developments (e.g. 3 du. **tas* < *ttas* < *d-tas*) that led to remodelling, and bases in *-ā* generalised this vowel in root presents, as in *yāti*, *yātas*, *yānti* (*yā* 'go, travel'). Moreover, there are some verbs with inherited invariant root presents, such as *ās*, *vas* 'have on, wear' (*vas-te*, *vas-āte*, *vas-ate*), *śī* 'lie, recline' (*śe-te*, *śay-āte*, *śe-rate*). Further, root presents of verbs in *-u* have *-au* instead of *-o* in alternation with *-u*; e.g. *stau-ti*, *stu-tas*, *stuv-anti* (*stu* 'praise'). There are also reduplicated stems, as in *juho-ti*, *juhu-tas*, *juhv-ati* (*hu* 'offer oblations'). In addition, ablauting present-imperfect stems are formed with suffixes and an infix. Thus, *śakno-/śaknu-* (*śak* 'be able'), *cino-/cinu-* (*ci* 'gather, heap'), *suno-/sunu-* (*su* 'press juice out of something') have a suffix *-no-/-nu-* (*-nv-* before vowels, *-nuv-* if the root ends in a consonant): *śaknoti*, *śaknutas*, *śaknuvanti*; *cinoti*, *cinutas*, *cinvanti*, *cinute*, *cinvāte*, *cinvate*; *sunoti*, *sunute*, etc. But *chi-na-d-/chi-n-d-* (*chinatti*, *chinttas*, *chindanti*; *chintte*, *chindāte*, *chindate*) shows an infix *-na-/-n-* added to *chid* 'cut'. Stems such as *pu-nā-/pu-nī-/pu-n-* 'purify' (*punāti*, *punītas*, *punanti*, *punīte*, *punāte*, *punate*), with short root vowels (contrast *pū-ta-* 'purified'), reflect an inherited formation with an infix added to a laryngeal base (Proto-Indo-European **-ne-H-/-n-H-*), but the types *krī-ṇā-*... 'buy' (*krīṇāti krīṇite* etc.), *badh-nā-*... 'tie up' (*badhnāti* etc.), with *-nā* etc. after a long vowel (cf. *krī-ta-* 'bought') or a consonant, show that this has been reanalysed as a suffix comparable to *-no-/-nu-*. Historical developments led to the creation of a stem *karo-/kuru-* (*karoti*, *kurutas*, *kurvanti*, *kurute*, *kurvāte*, *kurvate*)

from *kṛ*, in addition to the earlier *kṛṇo-/kṛṇu*, which allowed the abstraction of a suffix *-o/-u-*, as in *tano-/tanu-* (*tanoti*, *tanute* etc.), comparable to *śakno-/śaknu-*, from *tan* 'stretch', although originally this was the same suffix as in the type *śakno-/śaknu-*, only with bases in *-n* (*tano-/tanu-* < **tn̥-neu-/tn̥-nu-*).

Third person active and medio-passive imperative forms respectively have *-u*, *ām* instead of *-i*, *-e* of present indicatives; e.g. *as-tu*, *s-antu*; *ās-tām*, *ās-ātām*, *ās-atām*. However, second singular active imperatives of stems in *-a* have no overt ending: *bhav-a*, *dīv-ya*, *tud-a*. The same is true of the type *ci-nu*. However, if *-u* of the suffix *-nu-* follows a cluster, the imperative retains the ending *-hi*: *śaknuhi*; and this ending has a variant *-dhi* after *juhu-* and consonant-final stems: *juhudhi*, *chindhi* (< *chinddhi*). In addition, following consonant-final stems one has *-āna-* for presents with *-nā-*: *punīhi*, *krīṇīhi*, but *badhāna*. Second singular middle imperatives have a suffix *-sva*: *āssva*, *edhasva*, *cinuṣva*. First person imperative forms are historically subjunctives (see section 1.2): *bhavāni*, *bhavāva*, *bhavāma*; *edhai*, *edhāvahai*, *edhāmahai*. Other forms simply have secondary endings. In addition, there is an imperative with *-tāt* for both second and third singular, which, according to Pāṇini's description, was used in wishing someone well, as in *jīvatāt* 'may you/he live long'.

Stems in *-a* form optatives with *-ī-/-īy-*; other stems have optatives with *-yā-/-y-* in active forms and *-ī-/-īy-* in medio-passive forms. Optatives have the usual secondary endings except for active third plural *-us*, middle first singular *-a*, third plural *-ran*. For example: *bhaveyam*, *bhaves*, *bhavet*, *bhaveva*, *bhavetam*, *bhavetām*, *bhavema*, *bhaveta*, *bhaveyus*; *edheya*, *edhethās*, *edheta*, *edhevahi*, *edheyāthām*, *edheyātām*, *edhemahi*, *edhedhvam*, *edheran*; *syām*, *syās*, *syāt*, *syāva*, *syātam*, *syātām*, *syāma*, *syāta*, *syus* (*as* 'be'); *āsīya*, *āsīthās*, *āsīta*, *āsīvahi*, *āsīyāthām*, *asīyātām*, *āsimahi*, *āsīdhvam*, *āsīran*. Although synchronically the types *bhavet*, *edheta* are analysable as containing *-īy-/-ī-* (*-ey-* < *-a-īy-*, *-e-* < *-a-ī-*), these correspond to optatives elsewhere in Indo-European that point to **-oi-*. In addition, the use of *-yā-* in active and *-ī-* in medio-passive forms represents a redistribution of ablaut variants of an original single affix.

Aorists are either radical or formed with suffixes. Unreduplicated root aorists are rare in Classical Sanskrit as compared with earlier Vedic. Except for the third person singular passive aorist type *akāri* 'has been made' — which is freely formed to any verb, but is not necessarily to be analysed as a root aorist — only active forms of bases in *-ā* (e.g. *dā* 'give': *adāt*, *adātām*, *adus*) and of *bhū* 'be, become' (*abhūt*, *abhūtām*, *abhūvan*) regularly belong to this type, although some middle forms of root aorists have been incorporated into the sigmatic system. There are also stems in *-a*, such as *agama-* (*agamat*, *agamatām*, *agaman*: *gam* 'go'), *aghasa-* (*ghas* 'eat'), *aśaka-* (*śak* 'be able'). In addition, a reduplicated stem in *-a* regularly corresponds to a causative (see below) and supplies aorist forms to a few other verbs; e.g. *adudruva-* (*dru* 'run'). However, the productive Sanskrit aorist formation is

sigmatic, of four subtypes: *-s-*, *-iṣ*, *-siṣ-*, *-sa-*. The last developed from the middle of the *s*-aorist of *duh* (e.g. 1 sg. *adhukṣi*, 3 sg. du. pl. *adugdha*, *adhukṣātām*, *adhukṣata*), as can be seen from the earliest usage in Vedic, from the fact that *s*-forms are indeed incorporated into the *sa*-paradigm (e.g. mid. 1 sg. *adhukṣi*, 3 sg. *adugdha/adhukṣata*), and from the fact that this aorist is formed only with verbs that have penultimate *i*, *u*, *ṛ* and final consonants which give *-kṣ-* in combination with the *-s-* of the suffix. The *s*-aorist itself is characterised by particular variants of roots preceding the suffix. Verbs with *-ĭ*, *-ŭ*, *-ṛ* have alternants with *-ai*, *-au*, *-ār* before *-s-* in active forms, and verbs with *-ĭ*, *-ŭ* have variants with *-e*, *-o* in medio-passive forms; e.g. *ci*: *acaiṣīt*, *acaiṣṭām*, *acaiṣus*, *aceṣṭa*, *aceṣātām*, *aceṣata*; *hu*: *ahauṣīt*, *kṛ*: *akārṣīt* (but middle *akṛta akṛṣātām akṛṣata*). Verbs with medial vowels also have alternants with vṛddhi vowels in active forms, but they have medio-passives with *-a-*, *-i-*, *-u-*, *-ṛ-*; e.g. *pac* 'cook': *apākṣīt*, *chid*: *achaitsit*, *rudh* 'obstruct': *arautsit*, *mṛṣ* 'suffer, allow': *amārṣīt* versus *apakta*, *achitta*, *aruddha*, *amṛṣṭa*. Forms such as *akṛta*, *adita* (*dā* 'give') beside *akṛṣātām*, *adiṣātām* etc. and active *adāt* etc. reflect the incorporation of root aorist forms into the productive sigmatic system. The *iṣ*-aorist is probably best considered originally an *s*-formation to verbs with *-i* from a laryngeal, then spread well beyond these limits. This also has vṛddhi vowels in forms such as *apāvīt*, *apāviṣṭām*, *apāviṣus* (*pū*), but in general not for consonant-final bases; e.g. *div* 'gamble': *adevīt*. The *siṣ*-aorist, obviously a combination of *-s-* and *-iṣ-*, is of very limited compass, predominantly from verbs in *-ā*; e.g. *ayāsīt* (*yā*).

Although scholars disagree concerning the historical origins of the precative, the place of the forms in question within the Sanskrit system viewed synchronically is fairly clear. The active precative type *bhuyāt*, *bhuyāstām*, *bhuyāsus* 'may... be, prosper' is radical, and the middle type *edhiṣīṣṭa*, *edhiṣīyāstām*, *edhiṣīran* 'may... thrive' is sigmatic.

The semantically unmarked future of Sanskrit has a suffix *-(i)ṣya* after a root. In addition, there is a future used with reference to a time beyond the day of reference. In origin, this is a periphrastic formation (see section 1.2), but synchronically it cannot be treated as such in view of forms like *edhitāhe*, *edhitāsve*, *edhitāsmahe* (1 sg. du. pl. mid.), since *as* does not regularly have middle inflection. The future in *-(i)ṣya* (e.g., *bhaviṣyati*, *edhiṣyate*) is the basis for the Sanskrit conditional, of the type *abhaviṣyat*, *aidhiṣyata* — with augment and secondary endings — used in both the protasis and the apodosis of contrary-to-fact conditional sentences.

The Sanskrit perfect is generally characterised not only by particular endings but also by reduplication (see above). Yet one inherited perfect, which in Sanskrit functions as a present, lacks reduplication: *veda*, *vidatus*, *vidus* 'know(s)'. As can be seen, perfect stems show the same kind of grammatical alternation as found in present and aorist stems. However, for verbs of the structure C*a*C, in which *-a-* is flanked by single consonants the

first of which is not subject to modification in a reduplicated syllable, instead of -CC- preceded by a reduplicated syllable, one has CeC alone; e.g. *tan*: *tatāna*, *tenatus*, *tenus*; *śak*: *śaśāka*, *śekatus*, *śekus* (contrast *gam*: *jagāma*, *jagmatus*, *jagmus*). This represents the spread of a particular form from verbs like *yam* 'extend' (*yayāma*, *yematus* (< *ya-ym-*) ...), *sad* 'sit' (*sasada*, *sedatus* (< *sa-zd-*) ...). There is also a periphrastic perfect, which in Sanskrit has been extended to some primary verbs; e.g. *hu*: *juhavāñ cakāra* beside *juhāva*.

As can be seen from what has been said, it is not possible in Sanskrit to predict an aorist formation from the present-imperfect stem of a verb. There are instances where totally separate roots are used suppletively in different formations. Thus, *as* supplies only a present-imperfect stem; other forms are from *bhū* 'be, become': aorist *abhūt*, future *bhaviṣyati*, perfect *babhūva*, infinitive *bhavitum*, past participle *bhūta-* etc. Similarly: *han* 'strike, kill': aorist *avadhīt*, precative *vadhyāt*, *ad* 'eat': aorist *aghasat*, *i*: aorist *agāt*.

Derived verbs are deverbative or denominative. Causatives are formed with *-i-/-e-*; e.g. *kṛ*: *kār-i* 'have ... do, make' (*kār-ay-a-ti*, *kār-ay-ate*), *pac*: *pāc-i*, *chid*: *ched-i*, *yuj-* 'connect, yoke': *yoj-i*. Certain verbs have augmented variants before the causative suffix. For example, many verbs with *-ā* take the augment *-p*, as in *dāp-i* 'have ... give' (*dā*). The causative is also connected with a particular active aorist formation, a reduplicated *a*-aorist; e.g. *kār-i*: *acīkarat* etc. (but medio-passive *akārayita*, *akārayiṣātām*, etc.). Desideratives are formed with *-sa-*, which conditions reduplication; e.g. *kṛ*: *cikīrṣa-* (*cikīrṣati*, etc.). Desiderative forms alternate with phrases consisting of a verb meaning 'wish' and infinitives; e.g. *cikīrṣati* = *kartum icchati* '... wishes to do, make'. Intensives are formed with *-ya-*, which also conditions a particular type of reduplication; further, intensives have middle inflection; e.g. *kṛ*: *cekrīya-* (*cekrīyate*) 'do intensely, repeatedly', *chid*: *cechidya-*, *yuj*: *yoyujya-*, *pac*: *pāpacya-*. Derived verbs form periphrastic perfects, as in *gamayāñ cakāra*, *cekrīyāñ cakre*. Moreover, such deverbative formations can involve suppletion; e.g. *ad*: desiderative *jighatsa-*, *i*: *jigamiṣa-*. Denominatives are formed with several suffixes, principal among which is *-ya-*, and have a broad range of meanings. For example, *putrīyati* (*putrīya-*) corresponds to *putram icchati* '... desires a son', *putram ivācarati* '... behaves (*ācarati*) towards ... as though he were his son (*putram iva*)'; *śyenāyate* corresponds to *śyena ivācarati* 'behaves like a falcon (*śyena iva*)', *tapasyati* is equivalent to *tapaś carati* 'carries out (*carati*) ascetic acts (*tapas*).' Especially noteworthy in view of the later Indo-Aryan causative type in *-āv-e-* (see section 2.2 of the chapter on Indo-Aryan) is the denominative type *satyāpi-* (*satyāpayati*) 'say something is true (*satya*)', known already to Pāṇini, which involves *-āp-* and the suffix *-i-/-e-*.

2.2.4 Syntax

In major aspects of syntax Sanskrit is a fairly conservative Indo-European language, although it exhibits specifically Indic features. Examples given in the following sketch are based on Pāṇinian sources, reflecting usage that antedates classical literary works, but every construction illustrated has a counterpart in Vedic (see section 1.2) and literary texts of later times.

The seven cases of the nominal system excluding the vocative (section 2.2.2) are used with reference to various roles participants play in respect of what is signified by verbs in general or by particular verbs. Typical roles and case forms linked with them are illustrated by the following. In *devadattaḥ kaṭaṁ karoti* 'Devadatta is making (*karoti*) a mat (*kaṭam*)', *devadatto grāmaṁ gacchati* 'Devadatta is going (*gacchati*) to the village (*grāmam*)', the accusatives *kaṭam*, *grāmam* refer to objects, the latter specifically to a goal of movement. Such a goal is alternatively signified by a dative: *devadatto grāmāya gacchati*. In addition, an object can be designated by a genitive in construction with an agent noun; e.g. *sa kumbhānāṃ kartā* 'he (*sa*) (is) a maker (*kartā*) of pots (*kumbhānām*)'. In the passive sentence *devadattena kaṭaḥ kriyate* 'a mat is being made (*kriyate*) by Devadatta', the instrumental *devadattena* refers to an agent, as does the same form in *devadattena supyate* (section 2.2.3). The instrumental *dātreṇa* 'sickle' of *dātreṇa lunāti* '...cuts (*lunāti*) with a sickle', on the other hand, refers to a means of cutting. A dative can be used with references not only to a goal of movement but also to a desired object, in construction with *spṛh* 'yearn for': *puṣpebhyaḥ spṛhayati* '... yearns for flowers (*puṣbebhyas*)'. More generally, dative forms designate indirect objects, as in *māṇavakāya bhikṣāṁ dadāti* '... gives (*dadāti*) alms (*bhikṣām*) to the lad (*māṇavakāya*)'. Ablatives can be used to signify points of departure, as in *grāmād ā gacchati* '...is coming (*ā gacchati*) from the village', but they have other functions as well; for example, in *vṛkebhyo bibheti* '... is afraid (*bibheti*) of wolves', *vṛkebhyas* refers to wolves as sources of fear. Locative forms are used of loci where agents and objects are while they are involved in whatever a verb signifies; e.g. *devadattaḥ sthālyāṁ gṛha odanaṁ pacati* 'Devadatta is cooking (*pacati*) rice (*odanam*) in a pot (*sthālyām*) in the house (*gṛhe*)'.

There are also relations that do not directly involve verb meanings, so that syntactically one has nominals directly linked with each other. The typical case form for such relations is the genitive; e.g. *vṛkṣasya śākhā-* 'branch (*śākhā-*) of a/the tree (*vṛkṣasya*)' in *vṛkṣasya śākhāṁ paraśunā chinatti* '... is cutting a branch (*śākhām*) of the tree with an axe (*paraśunā*)'. Particular nominals, however, co-occur with other case forms. For example, *namo devebhyaḥ* '(let there be) homage (*namas*) to the gods' has the dative *devebhyas* in construction with *namas*. Moreover, pre- and postposed particles take part in such constructions: *sādhur devadatto mātaraṁ prati* 'Devadatta (is) good (*sadhus*) towards his mother (*mātaraṁ prati*)', *putreṇa sahāgataḥ* 'he came (*āgatas*) with his son (*putreṇa saha*)', *māṣān asmai tilebhyaḥ prati dadāti* '... gives (*dadāti*) this man (*asmai*) māṣa-beans (*māṣān*) in exchange for sesame seeds (*tilebhyaḥ prati*)', *ā pāṭaliputrād*

varṣati 'it is raining (*varṣati*) up to Pāṭaliputra (*ā pāṭaliputrāt*)', have the accusative *mātaram* linked to *prati*, the instrumental *putreṇa* connected to *saha*, and the ablatives *tilebhyas*, *pāṭaliputrāt* construed with *prati* and *ā*.

There are different kinds of complex sentences. Some involve related finite verb forms, others finite forms connected with particular nominal derivates, infinitival and participial. For example, optatives are used in conditional sentences such as *mriyeya ... na syās tvaṁ yadi me gatiḥ* 'I would die (*mriyeya*) if (*yadi*) you (*tvam*) were (*syās*) not (*na*) my (*me*) refuge (*gatis*)', but *edhān āhartum gacchati* '... is going (*gacchati*) in order to fetch (*āhartum*) firewood (*edhān*)' has *gacchati* linked to the infinitive *āhartum*, itself connected with the accusative *edhān*. There is an elliptical version of the second sentence type, with a dative referring to the direct object in question: *edhebhyo gacchati* '... is going for firewood'. Present participle forms occur in complex sentences such as *pacantaṁ devadattam paśyati* '... is watching (*paśyati*) Devadatta cook', in which *pacantam* 'cooking' agrees with *devadattam*, or *grāmaṁ gacchatā devadattena bhuktam* 'Devadatta ate on his way to the village', where the participial form *gacchatā* 'going' agrees with the agentive instrumental *devadattena*, both construed with *bhuktam* 'eaten'. In addition, Sanskrit has absolute constructions, the prevalent one being a locative absolute, as in *goṣu duhyamānāsu gataḥ* 'he left (*gatas*) while the cows were being milked': the present participle *duhyamānāsu* (loc. pl. f.) agrees with *goṣu* 'cows', both used absolutely. Where two or more verbs signify sequentially related acts or states, Sanskrit subordinates by using gerunds; e.g. *bhuktvā vrajati* '... eats before going out', with the gerund *bhuktvā* 'after eating', *piba niṣadya* (see section 1.2).

Examples cited illustrate the agreement features of Sanskrit. Finite verb forms — which themselves signal person and number differences — agree in person and number with nominals that function as grammatical subjects used in referring to agents or objects. Participial forms and other adjectivals, whether attributive or predicative, agree in gender and number with the nominals to which they are complements. The examples also illustrate the most common aspects of Sanskrit word order. What may be called the neutral word order in prose, where metrical constraints are not at play, generally has the verb in last position. However, a sentence does not necessarily have an overt verb: Sanskrit has nominal sentences, in which a third person present form of a verb meaning 'be' is not overtly expressed. There are few restrictions on word order that are strictly formal, but the position of certain particles is fixed: particles like *vai* 'as is known, truly', *ced* 'if ' occupy second position, as does *ca* 'and' used as a sentence connective. Similarly, the enclitic pronouns *mā*, *tvā* etc. (section 2.2.2) are excluded from sentence-initial position.

An aspect of overall sentence prosody is worth noting in this context. A sentence-internal vocative generally has no high-pitched vowel. Under certain conditions, however, the vowels of an utterance are all pronounced

monotone, except for the last vowel, which is then not only high-pitched but also prolated. For example, in *ā gaccha bho māṇavaka devadatta* 'come along (*ā gaccha*), Devadatta my boy (*bho māṇavaka devadatta*)', used in calling Devadatta from afar, all the vowels up to the *-a* of the vocative *devadatta* are uttered without pitch variations, but this last vowel is prolated and udātta.

Bibliography

Burrow (1965) is a summary of the prehistory and history of Sanskrit, including Vedic, with references to Middle Indo-Aryan; somewhat personal views are given in places, but the work remains valuable. For a good summary of views on the dialects of Old Indo-Aryan, with discussion of theories proposed and references, see Emeneau (1966).

The standard reference grammar is Whitney (1889). Renou (1956) is an insightful summary of the grammar, vocabulary and style of different stages of Sanskrit, including Vedic, with text selections and translations. Wackernagel (1896–) is the most thorough reference grammar of Sanskrit, but remains incomplete: the published volumes are: I (*Lautlehre*), reissued with a new 'Introduction générale' by L. Renou and 'Nachträge' by A. Debrunner (1957); II, 1 (*Einleitung zur Wortlehre, Nominalkomposition*), 2nd ed. with 'Nachträge' by A. Debrunner (1957); II, 2 (*Die Nominalsuffixe*), by A. Debrunner (1954); III (*Nominalflexion – Zahlwort – Pronomen*) (1930); there is also a *Register zur altindischen Grammatik von J. Wackernagel und A. Debrunner* by R. Hauschild (1964).

References

Burrow, T. 1965. *The Sanskrit Language*, 2nd ed. (Faber and Faber, London)

Emeneau, M.B. 1966. 'The Dialects of Old Indo-Aryan', in H. Birnbaum and J. Puhvel (eds.), *Ancient Indo-European Dialects* (University of California Press, Berkeley and Los Angeles), pp. 123–38.

Renou, L. 1956. *Histoire de la langue sanskrite* (IAC, Paris)

Wackernagel, J. 1896–. *Altindische Grammatik* (Vandenhoek and Ruprecht, Göttingen)

Whitney, W.D. 1889. *Sanskrit Grammar, Including Both the Classical Language and the Older Dialects, of Veda and Brahmana*, 2nd ed. (Harvard University Press, Cambridge, Mass.)

22 Hindi-Urdu

Yamuna Kachru

1. Introduction

Hindi is a New Indo-Aryan language spoken in the north of India. It belongs to the Indo-Iranian branch of the Indo-European family of languages. It is spoken by more than two hundred million people either as a first or second language in India, and by peoples of Indian origin in Trinidad, Guyana, Fiji, Mauritius, South Africa and other countries. Along with English, it is the official language of India. In addition, it is the state language of Bihar, Haryana, Himachal Pradesh, Madhya Pradesh, Rajasthan and Uttar Pradesh.

Urdu, a language closely related to Hindi, is spoken by twenty-three million people in India and approximately eight million people in Pakistan as a mother tongue. It is the official language of Pakistan and the state language of the state of Jammu and Kashmir in India.

It is difficult to date the beginnings of the New Indo-Aryan languages of India. Scholars generally agree that the development of Indo-Aryan languages of India took place in three stages. The Old Indo-Aryan stage is said to extend from 1500 BC to approximately 600 BC. The Middle Indo-Aryan stage spans the centuries between 600 BC and AD 1000. The Middle Indo-Aryan stage is further subdivided into an early Middle Indo-Aryan stage (600–200 BC), a transitional stage (200 BC–AD 200), a second Middle Indo-Aryan stage (AD 200–600), and a late Middle Indo-Aryan stage (AD 600–1000). The period between AD 1000–1200/1300 is designated the Old New Indo-Aryan stage because it is at this stage that the changes that began at the Middle Indo-Aryan stage became established and the New Indo-Aryan languages such as Hindi, Bengali, Marathi etc. assumed distinct identities.

Before proceeding with a description of Hindi-Urdu, it may be useful to sketch briefly the sociolinguistic situation of Hindi-Urdu in the Indian subcontinent (Rai 1984).

The name Hindi is not Indian in origin; it is believed to have been used by the Persians to denote the peoples and languages of India (Verma 1933). Hindi as a language is said to have emerged from the patois of the market

place and army camps during the period of repeated Islamic invasions and establishment of Muslim rule in the north of India between the eighth and tenth centuries AD. The speech of the areas around Delhi, known as *khaṛī bolī*, was adopted by the Afghans, Persians and Turks as a common language of interaction with the local population. In time, it developed a variety called *urdū* (from Turkish *ordu* 'camp'). This variety, naturally, had a preponderance of borrowings from Arabic and Persian. Consequently, it was also known as *rextā* 'mixed language'. The speech of the indigenous population, though influenced by Arabic and Persian, remained relatively free from large-scale borrowings from these foreign languages. In time, as Urdu gained some patronage at Muslim courts and developed into a literary language, the variety used by the general population gradually replaced Sanskrit, literary Prakrits and Apabhraṁśas as the literary language of the midlands (*madhyadeśa*). This latter variety looked to Sanskrit for linguistic borrowings and Sanskrit, Prakrits and Apabhraṁśas for literary conventions. It is this variety that became known as Hindi. Thus, both Hindi and Urdu have their origins in the *khaṛī bolī* speech of Delhi and its environs although they are written in two different scripts (Urdu in Perso-Arabic and Hindi in Devanāgarī). The two languages differ in minor ways in their sound system, morphology and syntax. These differences are pointed out at appropriate places below.

Hindi and Urdu have a common form known as Hindustani which is essentially a colloquial language (Verma 1933). This was the variety that was adopted by Mahatma Gandhi and the Indian National Congress as a symbol of national identity during the struggle for freedom. It, however, never became a language of literature and high culture (see Bhatia 1987 for an account of the Hindi-Urdu-Hindustani controversy in the late nineteenth and early twentieth centuries).

Both Urdu and Hindi have been in use as literary languages since the twelfth century. The development of prose, however, begins only in the eighteenth century under the influence of English, which marks the emergence of Hindi and Urdu as fully-fledged literary languages.

2 Phonology

The segmental phonemes of Hindi-Urdu are listed in table 22.1. The phonemes that occur only in the highly Sanskritised or highly Persianised varieties are given in parentheses. The two noteworthy features of the inventory of consonant phonemes are the following: Hindi-Urdu still retains the original Indo-European distinction between aspirated and unaspirated voiced plosives (cf. Indo-European **ghr̥dho* and Hindi *ghər* 'house'). It retains the distinction between aspirated and unaspirated voiceless plosives that emerged in Indo-Aryan, i.e. the distinction between *kal* 'time' and *khal*

'skin'. Another Indo-Aryan feature, that of retroflexion, is also retained in Hindi-Urdu, cf. *tota* 'parrot' and *ṭoṭa* 'lack'. These two features, i.e. those of aspiration and retroflexion, are mainly responsible for why Hindi-Urdu sounds so different from its European cousins.

Table 22.1: Phonemes of Hindi-Urdu

Vowels

	Front	*Centre*	*Back*
High	i		u
	ɪ		ʊ
Mid High	e		o
Mid Low	ɛ	ə	ɔ
Low		a	

Consonants

			Labial	*Dental*	*Retro-flex*	*Alveo-Palatal*	*Velar*	*Back Velar*
Stop	vls.	unasp.	p	t	ṭ	č	k	(q)
		asp.	ph	th	ṭh	čh	kh	
	vd.	unasp.	b	d	ḍ	ǰ	g	
		asp.	bh	dh	ḍh	ǰh	gh	
Nasal			m	n	(ṇ)	(ñ)	(ŋ)	
Flap	vd.	unasp.			ɾ	r		
		asp.			ɾh			
Lateral						l		
Fricative	vls.		(f)	s	(ṣ)	š	(x)	
	vd.			(z)		(ž)	(ɣ)	
Semi-vowels			w (v)			y		

Note: Oral and nasal vowels contrast, e.g. *ak* 'a plant' and *āk* 'draw, sketch'; hence, nasalisation is distinctive. Short and long consonants contrast, e.g. *pəta* 'address', *pətta* 'leaf'; hence, length is distinctive.

The contrast between aspirated and unaspirated consonants is maintained in all positions, initial, medial and final. The distinction between tense *i* and lax *ɪ* and tense *u* and lax *ʊ*, however, is lost in the final position except in very careful and formal speech in the highly Sanskritised variety.

Stress is not distinctive in Hindi-Urdu; words are not distinguished on the basis of stress alone. For instance, a word such as *kəla* 'art', whether stressed as *'kəla* or *kə'la*, means the same. The tense vowels are phonetically long and in pronunciation the vowel quality as well as length is maintained irrespective of the position of the vowel or stress in the word. For instance, the word *muskərahəṭ* 'smile' can either be stressed as *'muskərahəṭ* or *muskə'rahəṭ*, in either case, the vowel quality and length in the syllable *-ra-*

remains unaffected. Words such as *ǰamata* 'son-in-law' are pronounced with three successive long vowels although only the first or the second syllable is stressed. Stressing and destressing of syllables is tied to syllable weight in Hindi-Urdu. Syllables are classified as one of the three measures of weight: light (syllables ending in a lax, short vowel), medium (syllables ending in a tense, long vowel or in a lax, short vowel followed by a consonant) and heavy (others). Where one syllable in a word is of greater weight than others, the tendency is to place the word stress on it. Where more than one syllable is of maximum weight in the word (i.e. there is a succession of medium or heavy syllables), usually the last but one bears the word stress. This stress pattern creates the impression of the staccato rhythm that speakers of English notice about Hindi-Urdu.

The predominant pattern of penultimate stress in Hindi-Urdu is inherited from an earlier stage of Indo-Aryan, i.e. the Middle Indo-Aryan stage. Old Indo-Aryan had phonemic accent of the pitch variety and there is evidence for three pitches in Vedic: *udātta* 'high, raised', *anudātta* 'low, unraised' and *svarita* 'high falling, falling' (see section 2.1 of the chapter on Sanskrit). At a later stage of Old Indo-Aryan, Classical Sanskrit does not record accent. By late Old Indo-Aryan, pitch accent seems to have given way to stress accent. There are different opinions about stress accent in Middle Indo-Aryan. It is generally believed that stress occurred on the penultimate syllable of the word, if long, or on the nearest preceding syllable if the penultimate was not long; in words with all short syllables, stress occured on the initial syllable.

Syllable boundaries in Hindi-Urdu words fall as follows: between successive vowels, e.g. *pa-e* 'legs', *a-ɪ-e* 'come' (hon.), *nə-i* 'new' (f.), *so-ɪ-e* 'sleep' (hon.); between vowels and following consonants, e.g. *ro-na* 'to cry', *pə-ta* 'address', *ū-ča* 'tall, high'; between consonants, e.g. *səṛ-kē* 'roads', *pət-la* 'thin', *hɪn-di* 'Hindi language'.

As has already been said, Hindi is written in the Devanāgarī script, which is the script used by Sanskrit, Marathi and Nepali also. On the basis of the evidence obtained from the ancient inscriptions, it is clear that Devanāgarī is a descendant of the Brāhmī script. Brāhmī was well established in India some time before 500 BC. Despite some controversy regarding the origin of the Brāhmī script, it is generally believed that its sources lie in the same Semitic script which later developed into the Arabic, Hebrew, Greek, Latin scripts etc. The scripts used for the New Indo-Aryan and the Dravidian languages of India are believed to have developed from the northern and southern varieties of Brāhmī.

There are minor differences between the scripts used for Hindi, Sanskrit, Marathi and Nepali. For instance, Hindi does not have the retroflex lateral ळ or the retroflex vowels ऋ, ॠ and ऌ. It uses the retroflex vowel symbol ऋ and the symbol for weak aspiration ः only in words borrowed from Sanskrit. Although written as ऋ, the vowel is pronounced as a combination of *r* and *ɪ*.

In general, there is a fairly regular correspondence between the script and

the pronunciation. The one notable exception is the pronunciation of the inherent vowel *ə*. The Devanāgarī script is syllabic in that every consonant symbol represents the consonant plus the inherent vowel *ə*, thus, the symbol क represents the sound *k* plus *ə*, or *kə*. Vowels are represented differently according to whether they comprise entire syllables or are parts of syllables, i.e. are immediately preceded by a consonant: thus, the symbol ई represents the syllable *i*, but in the syllable *ki*, it has the shape ि which is adjoined to the symbol for *k*, resulting in की. Even though each consonant symbol represents a consonant plus the inherent vowel, a word written as कल, i.e. *kələ*, is not pronounced as *kələ*, it is pronounced as *kəl* 'yesterday, tomorrow'. That is, all the final inherent vowels are dropped in pronunciation. The rules regarding the realisation of the inherent vowel in pronunciation are as follows; in two or three syllable words, the penultimate inherent vowel is pronounced when the final one is dropped, and in words of four syllables, both the final and the antepenultimate inherent vowels are dropped while the others are pronounced. Thus, *səməǰhə* is pronounced as *səməǰh* 'understanding', *mehənətə* is pronounced as *mehnət* 'hard work'. These general principles, however, do not apply to words containing medial *h*, loanwords, compounds and words formed with derivational suffixes. For instance, *səməǰh* with the inflectional suffix of perfective *-a* is pronounced as *səmǰha* 'understood', but with the derivational agentive suffix *-dar* is pronounced *səməǰhdar* 'sensible' (see Ohala (1983) for details of *ə*-deletion).

Although most derivational and inflectional morphology of Hindi is affixal in nature (i.e. Hindi mostly utilises prefixes and suffixes), there are remnants of the morphophonemic ablaut alternation of vowels of the *guṇa* and *vr̥ddhi* type (see pages 43–4) in a substantial number of verbal roots and nominal compounds in Hindi. These are the most frequent and regular of vowel changes for derivation as well as inflection in Sanskrit. A *guṇa* vowel differs from a simple vowel by a prefixed *a*-element which is combined with the other according to the usual rules; a *vr̥ddhi* vowel, by the further prefixation to a *guṇa* vowel. *a* is its own *guṇa* and *ā* remains unchanged for both *guṇa* and *vr̥ddhi*. The series of corresponding degrees is as follows (Kellogg 1875):

Simple vowels:	a	ā	i	ī	u	ū	r̥	l̥
guṇa vowels:	a	ā	e		o		ar	al
vr̥ddhi vowels:		ā	ai		au		ār	

The *guṇa* increment is an Indo-European phenomenon, the *vr̥ddhi* increment is specifically Indian in origin. These processes are still utilised to some extent in coining new compounds of borrowings from Sanskrit for modernising Hindi. Some examples of the verbal roots that exemplify these processes are pairs such as *khul* 'open' (intr.) and *khol* 'open' (tr.); *kəṭ* 'cut' (intr.) and *kaṭ* 'cut' (tr.), *dɪkh* 'be visible' and *dekh* 'see'; and some examples of nominal compounds are *pərəmə* + *išvərə* = *pərəmešvər* 'Supreme God'; *məha* + *išə* = *məheš* 'Great God' (a name of Śiva); *sədə* + *evə* = *sədɛv* 'always'. Some examples of modern vocabulary coined on the same principles are *sərvə* + *udəyə* = *sərvodəy* 'universal welfare', *mətə* + *ɛkyə* = *mətɛky* 'unanimity of opinion', *šʊbhə* + *ɪčču* = *šʊbheččʊ* 'well wisher'.

Table 22.2 gives the Devanāgarī script as used for Hindi:

Table 22.2: Chart of Devanāgarī Alphabet

Vowels
Independent

अ	आ	इ	ई	उ	ऊ	ऋ
ə	a	ɪ	i	ʊ	u	rɪ
ए	ऐ	ओ	औ	अं	अः	
e	ɛ	o	ɔ	əm	əh	

Following Consonant

ा	ि	ी	ु	ू	े	ै	ो	ौ	ं	ः
a	ɪ	i	ʊ	u	e	ɛ	o	ɔ	əm	əh

Consonants

क	ख	ग	घ	ङ				
kə	khə	gə	ghə	ŋə				
च	छ	ज	झ	ञ				
čə	čhə	ǰə	ǰhə	ñə				
ट	ठ	ड	ढ	ण				
ṭə	ṭhə	ḍə	ḍhə	ṇə				
त	थ	द	ध	न				
tə	thə	də	dhə	nə				
प	फ	ब	भ	म				
pə	phə	bə	bhə	mə				
य	र	ल	व	श	ष	स	ह	
yə	rə	lə	və	šə	ṣə	sə	hə	
क़	ख़	ग़	ज़	फ़				
qə	xə	ɣə	zə	fə				

To the extent that it shares a basic vocabulary with Hindi, the *guṇa* and *vr̥ddhi* phenomena are applicable to Urdu as well. The Urdu writing system, however, is based on the Perso-Arabic script. As is clear from table 22.3, the script lacks adequate vowel symbols but has an overabundance of consonant symbols for the language. Table 22.3 lists the independent forms only (see also the discussion of script in the chapters on Arabic and Persian).

Table 22.3: The Urdu Alphabet

Letter	*Pronunciation*	*Urdu Name*
ا	a*	əlyf
ب	b	be
پ	p	pe
ت	t	te
ٹ	ṭ	ṭe
ث	s	se
ج	ǰ	ǰim
چ	č	če
ح	h	he [/bəṛi he/]
خ	x	xe
د	d	dal
ڈ	ḍ	ḍal
ذ	z	zal
ر	r	re
ڑ	ṛ	ṛe
ز	z	ze
ژ	ž	že
س	s	sin
ش	š	šin
ص	s	swad
ض	z	zwad
ط	t	to, toe
ظ	z	zo, zoe
ع	*	əyn
غ	ɣ	ɣəyn
ف	f	fe
ق	q	qaf
ک	k	kaf
گ	g	gaf
ل	l	lam
م	m	mim
ن	n	nun
و	v	vao
ہ	h	he [/choṭi he/]
ی	y	ye

Note: əlyf is pronounced as *ā* following a consonant; əyn is either not pronounced at all or given the value of *a* or *ā* following a consonant. It is pronounced as a glottal stop only in High Urdu.

3 Morphology

A brief description of Hindi-Urdu nominal and verbal morphology follows (for a detailed discussion of derivational and inflectional morphology, see McGregor (1972), Sharma (1958) and Bailey (1956)).

3.1 Nominal

Forms of Hindi-Urdu nouns undergo changes in order to indicate number, gender and case. There are two numbers, singular and plural; two genders, masculine and feminine; and three cases, direct, oblique and vocative. Nouns are declined differently according to the gender class and the phonological property of the final segment in the word. Given here are paradigms of the major classes of masculine and feminine nouns.

Paradigm of Masculine Nouns Ending in *-a*

	Sg.	*Pl.*
Dir.	ləṛka 'boy'	ləṛke
Obl.	ləṛke	ləṛkõ
Voc.	ləṛke	ləṛko
Ending in *-i*		
Dir.	mali 'gardener'	mali
Obl.	mali	malıyõ
Voc.	mali	malıyo
Ending in *-u*		
Dir.	saṛhu 'wife's sister's husband'	saṛhu
Obl.	saṛhu	saṛhuõ
Voc.	saṛhu	saṛhuo
Ending in a consonant		
Dir.	nɔkər 'servant'	nɔkər
Obl.	nɔkər	nɔkərõ
Voc.	nɔkər	nɔkəro

Certain masculine nouns ending in *-a* such as *raǰa* 'king' and kinship terms such as *pıta* 'father', *čača* 'father's younger brother', *mama* 'mother's brother' are exceptions in that they do not change for direct plural and oblique singular in modern standard Hindi.

Paradigm of Feminine Nouns Ending in *-i*

	Sg.	*Pl.*
Dir.	ləṛki 'girl'	ləṛkıyā
Obl.	ləṛki	ləṛkıyõ
Voc.	ləṛki	ləṛkıyo
Ending in *-a*		
Dir.	mata 'mother'	mataẽ
Obl.	mata	mataõ
Voc.	mata	matao
Ending in *-u*		
Dir.	bəhu 'daughter-in-law'	bəhuẽ
Obl.	bəhu	bəhuõ
Voc.	bəhu	bəhuo

Ending in a consonant		
Dir.	bəhən 'sister'	bəhnẽ
Obl.	bəhən	bəhnõ
Voc.	bəhən	bəhno

In Perso-Arabic borrowings, High Urdu keeps the Perso-Arabic plural markers, e.g. *kaɣəz* 'paper': *kaɣzat* 'papers'.

The oblique case forms are used whenever a noun is followed by a postposition, e.g. *ləṛke ko* 'to the boy', *ghərõ mẽ* 'in the houses', *ləṛkɪyõ ke sath* 'with the girls' etc.

The adjectives occur before the noun and agree with their head noun in number, gender and case. They do not, however, exhibit the full range of forms. This can be seen in the paradigm of *əččhA* 'good' (*A* is a cover symbol for the various inflections).

əččhA 'good'

	Masculine		*Feminine*	
	Sg.	Pl.	Sg.	Pl.
Dir.	əčča	əčče	əčči	əčči
Obl.	əčče	əčče	əčči	əčči
Voc.	əčče	əčče	əčči	əčči

The adjectives that end in a consonant, e.g. *sʊndər* 'beautiful', and in a vowel other than *-a*, e.g. *nəkli* 'false, artificial', are invariant, e.g. *sʊndər ləṛka/ləṛki* 'handsome boy/beautiful girl', *nəkli dãt* (m.)/*bãh* (f.) 'artificial teeth/arm'.

The main postpositions that indicate case relations such as accusative, dative, instrumental etc. are the following: *ne* 'agentive, marker of a transitive subject in the perfective', *ko* 'accusative/dative', *se* 'instrumental/ablative/comitative', *mẽ*, *pər* 'locative', *kA* 'possessive/genitive', and *ke lɪye* 'benefactive'. There are several other postpositions that indicate location, direction, etc. such as *ke pas* 'near', *ki or* 'toward', *ke samne* 'in front of', *ke pĭche* 'behind', *ke bahər* 'out (of)', *ke əndər* 'inside', *ke par* 'across', *ke bɪna* 'without', *ke sath* 'with' and *ke hath/dvara* 'through'.

The pronouns have more case forms than the nouns, as is clear from the following paradigm:

	1st		*2nd*		*3rd*	
	Sg.	Pl.	Sg.	Pl.	Sg.	Pl.
Dir.	mẽ	həm	tu	tʊm	yəh/vəh	ye/ve
Obl.	mʊjh	həm	tʊjh	tʊm	ɪs/ʊs	ɪn/ʊn
Poss.	merA	həmarA	terA	tʊmharA	ɪs/ʊs kA	ɪn/ʊn kA

The third person pronominal forms are the same as the proximate and remote demonstratives, *yəh* 'this' and *vəh* 'that', and their inflected forms.

The possessive form of the pronouns behaves like an adjective and agrees with the possessed noun in number, gender and case, e.g. *mere beṭe ko* 'to my son', *tʊmhari kɪtabõ mẽ* 'in your books', *ʊnki bəhnõ ke sath* 'with their sisters' etc. The oblique forms are used with the postpositions except that the first and second person pronouns are used in their direct case forms with the agentive postposition *ne*. The third person plural pronouns have special combined forms when they are followed by the agentive postposition, e.g. *ɪn + ne = ɪnhõne* and *ʊn + ne = ʊnhõne*. All the pronouns listed above have special contracted forms when followed by the accusative/dative postposition, e.g. *mʊjh + ko = mʊjhe*, *tʊjh + ko = tʊjhe*, *ɪs/ʊs + ko = ɪse/ʊse*, *həm + ko = həmẽ*, *tʊm + ko = tʊmhẽ*, *ɪn/ʊn + ko = ɪnhẽ/ʊnhẽ*.

In addition to the pronouns listed above, Hindi-Urdu has a second person honorific pronoun *ap* which is used with both singular and plural reference for both male and female addressees. The honorific pronoun has the same form in all numbers and cases, i.e. it is invariant. The possessive is formed by adding the postposition *kA* to *ap*. To make the plural reference clear, the item *səb* 'all' or *log* 'people' may be added to the form *ap*, e.g. *ap səb/log*.

Hindi-Urdu also has a reflexive pronoun *ap* 'self ' which has an oblique form *əpne* and a possessive form *əpnA*. The form *ap* is used for all persons. There is a reduplicated form of *ap*, i.e. *əpne ap*, which is also used as the reflexive pronoun in Hindi-Urdu, e.g. *ram ne əpne ko/əpne ap ko šiše mẽ dekha* 'Ram looked at himself in the mirror'.

The two interrogative pronouns, *kən* and *kya* are used for human and non-human respectively. The oblique forms of these pronouns are *kɪs* in the singular and *kɪn* in the plural. The possessive is formed by adding the possessive postposition *kA* to the oblique. Similar to the third person pronouns, these pronouns also have combined forms such as *kɪnhõne*, *kɪse* and *kɪnhẽ*.

The devices of reduplication and partial reduplication or echo-compounding are used for expressing various meanings. For instance, reduplication of adjectives has either an intensive or a distributive meaning, e.g. *lal-lal saṛi* 'very red saree', *taza-taza dudh* 'very fresh milk', *kale-kale bal* 'jet-black hair', *ũče-ũče pəhaṛ* 'tall mountains', etc. Echo-compounding of adjectives, nouns and verbs has the meaning 'and the like', e.g. *sʊndər-vʊndər* 'pretty and such', *čay-vay* 'tea and other such things', *mɪlna-vɪlna* 'meeting and other such things' etc. The echo-compounding usually tones down the meaning of the adjective; it, however, adds to the meaning of other word classes. For instance, *čay-vay* means not only tea but snacks that go with tea, *pəṛhna-vəṛhna* means not only reading but other activities that go with studying.

In addition to reduplication and echo-compounding, another device used extensively is that of compounding two words with related meanings, e.g. *hãsi-xuši* 'laughter and happiness' (pleasant state or occasion), *dʊkh-taklif* 'sorrow and pain' (state full of sorrow), *šadi-byah* 'wedding' etc. Note that in

all these examples, one item is from Indic sources, the other from Perso-Arabic sources. This is extremely common, though not absolutely obligatory.

In Hindi-Urdu, the possessor normally precedes the possessed and the possessive postposition *kA* agrees with the possessed in number, gender and case, e.g. *ləṛke ki kɪtab* 'the boy's book', *ləṛke ke sɪr pər* 'on the boy's head' etc. High-Urdu has an alternative construction where the possessed precedes the possessor following the convention of the ezafe-construction in Persian (see page 532), e.g. *šer-e-kəšmir* 'the lion of Kashmir', *qəvaɪd-e-ʊrdu* 'grammar of Urdu', etc.

3.2 Verbal

Two most noticeable things about Hindi-Urdu verbs are their occurrence in morphologically related sets and in series. The first phenomenon is known as causal verbs and the second as compound verbs. Whereas the causative is inherited from Old Indo-Aryan, the development of compound verbs in New Indo-Aryan is recent — it became frequent only in the period between AD 600 and 1000.

Some examples of causal verbs can be seen in the chart given here.

Causal Verbs

Intr.	*Tr.*	*Dbl. tr.*	*Caus.*
ʊṭh 'rise'	ʊṭha 'raise'	–	ʊṭhva 'cause to rise/raise'
kəṭ 'be cut'	kaṭ 'cut'	–	kəṭva 'cause to (be) cut'
–	sʊn 'hear'	sʊna 'recite/narrate'	sʊnva 'cause to hear/narrate'
–	kha 'eat'	khɪla 'feed'	khɪlva 'cause to eat/feed'

Examples of compound verbs are *gɪr ǰana* 'fall go = fall down', *kha lena* 'eat take = eat up', *pəṛh lena* 'read take = read to oneself ', *pəṛh dena* 'read give = read out loud to someone'.

Hindi-Urdu verbs occur in the following forms: root, e.g. *kha* 'eat', *a* 'come', imperfect stem, e.g. *khatA*, *atA*, perfect stem, e.g. *khayA*, *ayA*, and infinitive, *khanA*, *anA*. The stems behave like adjectives in that they agree with some noun in the sentence in number and gender. The imperfect and perfect participles, which are made up of the imperfect and perfect stems followed by the perfect stem of the verb *ho* 'be', i.e. *hʊA*, agree in case also. This means that the stem final *-A* changes to *-e* or *-i* for agreement. Whereas the imperfect and perfect aspectual distinction is expressed by suffixation, the continuous aspect is indicated by an independent lexical item, *rəhA*. This marker follows the root and behaves like the imperfect and perfect stems with regard to gender and number agreement.

The tense distinction of present versus past is expressed with the forms of the auxiliary verb, the present auxiliary *hE* and the past auxiliary *thA*. These are the present and past forms of the stative verb *honA* 'be'. As in all Indo-

European languages, the verb 'be' is irregular in Hindi. It has the following forms: root *ho*, imperfect stem *hotA*, perfect stem *hʊA*, infinitive *honA*, stative present *hE*, stative past *thA*. The stem-final *-A* changes to *-e*, *-i* or *-ĩ* for number and gender agreement and the final *-E* changes to various vowels to indicate person, number and gender agreement. The forms of the verb *honA* in stative present are as follows: 1st person sg. *hũ*, 2nd and 3rd person sg. *hɛ̃*, 2nd person pl. *ho*, and 1st and 3rd person pl. and 2nd hon. *hɛ̃*.

In addition to tense and aspect distinctions, the verbal forms express mood distinctions as well. There is no distinction made between indicative and interrogative, i.e. in assertions as well as questions, the verbal forms are made up of the stems and auxiliaries described above. Historically, Old Indo-Aryan did not make a distinction between these two moods either. The moods in Old Indo-Aryan were indicative, imperative, optative and subjunctive. In Hindi-Urdu, the optative forms are made up of the root and the following suffixes: 1st person sg. *-ũ*, 2nd and 3rd person sg. *-e*, 1st and 3rd pl. and 2nd honorific *-ẽ*, and 2nd pl. *-o*. The future tense is formed by adding the suffix *-gA* to the optative forms, e.g. *ǰa-ũ-ga* 'I (m.) will go', *ǰaogi* 'you (f.) will go' etc. The following are the imperative forms: root form of the verb (intimate or rude), 2nd pl. optative (familiar), root with the suffix *-ɪye* (honorific, polite), root with the suffix *-ɪye* followed by the suffix *-ga* (remote, therefore, extra polite) and the infinitive form of the verb (remote imperative, therefore even when used with second plural, polite). Thus, the imperative forms of the verb *kha* are *(tu) kha* 'you (intimate) eat', *tʊm khao* 'you (familiar) eat', *(ap) khaɪye* 'you (honorific) eat', *(ap) khaɪyega* 'you (honorific) please eat (perhaps later?)', *(tʊm) khana* 'you (familiar, polite) eat' or 'you (familiar) eat (perhaps later?)'.

The paradigm of the verb *ghumna* 'to take a walk' illustrates the full range of the forms discussed above.

Paradigm of Verb Forms

Root: ghum 'take a walk'
Imperfect stem: ghumtA
Perfect stem: ghumA
Infinitive: ghumnA
Optative: ghumũ (1st sg.), ghumo (2nd pl.), ghume (2nd and 3rd sg.), ghumẽ (1st and 3rd pl., 2nd honorific)
Imperative: ghum (2nd sg., intimate/rude), ghumo (2nd pl., familiar), ghumɪye (2nd honorific, polite), ghumɪyega (2nd honorific, extra polite)

Future

	1st		*2nd*		*3rd*	
	M.	F.	M.	F.	M.	F.
Sg.	ghumunga	ghumungi	ghumega	ghumegi	ghumega	ghumegi
Pl.	ghumenge	ghumengi	ghumoge	ghumogi	ghumenge	ghumengi
Hon.	–	–	ghumenge	ghumengi	ghumenge	ghumengi

Present imperfect

		Sg.	*Pl.*	*Hon.*
1st	M.	ghumta hũ	ghumte hẽ	–
	F.	ghumti hũ	ghumti hẽ	–
2nd	M.	ghumta hɛ	ghumte ho	ghumte hẽ
	F.	ghumti hɛ	ghumti ho	ghumti hẽ
3rd	M.	ghumta hɛ	ghumte hẽ	ghumte hẽ
	F.	ghumti hɛ	ghumti hẽ	ghumti hẽ

Past imperfect: ghumta tha, ghumte the, ghumti thi, ghumti thī, etc.
Present perfect: ghuma hũ, ghumi hũ, etc.
Past perfect: ghuma tha, ghumi thi, etc.
Present continuous: ghum rəha hũ, ghum rəhi hũ, etc.
Past continuous: ghum rəha tha, ghum rəhi thi, etc.

In general, Urdu speakers use the masculine plural form as undifferentiated for gender in the first person, e.g. *həm kəl kəlkətte ǰa rəhe hẽ* 'We (m./f.) are going to Calcutta tomorrow.'

The contingent, past contingent and presumptive tenses are formed with the imperfect and perfect stems and the continuous form followed by the auxiliaries *ho* 'contingent', *hotA* 'past contingent', and *hogA* 'presumptive'. Roughly, these three are translatable into English as follows: *ata ho* '(he) may be coming', *aya ho* '(he) may have come', *ata hota* 'had (he) been coming', *aya hota* 'had (he) come', *ata hoga* '(he) must be coming', *aya hoga* '(he) must have come'.

Hindi-Urdu verbs are very regular, which means that once we know the infinitive form of the verb, we can isolate the root and derive the imperfect and perfect stems by suffixing *-tA* and *-A* respectively. Thus, from *hãsna* 'laugh', we get the imperfect stem *hãstA* and perfect stem *hãsA*. Note that when the root ends in a vowel and the perfect stem-forming suffix *-A* is added to it, a semi-vowel is inserted to separate the two vowels. If the root ends in *-i*, *-a* or *-o*, a *-y-* is inserted, if the root ends in *-u*, a *-v-* is inserted, e.g. *kha* + *-A* = *khaya* 'ate (m.)', *ro* + *-A* = *roya* 'cried (m.)', *pi* + *-A* = *pɪya* 'drank (m.)', *čhu* + *-A* = *čhʊva* 'touched (m.)'.

One verb, *čahɪye*, is completely irregular in that it has only this form. It takes a dative subject and means 'to need' or 'want'. The following have irregular perfect stems: *kər* 'do' – *kɪya*, *le* 'take' – *lɪya*, *de* 'give' – *dɪya*, *ǰa* 'go' – *gəya*. The following have irregular polite imperative forms: *kər* 'do' = *kiǰɪye*, *le* 'take' = *liǰɪye*, *de* 'give' = *diǰɪye*, *pi* 'drink' = *piǰɪye*.

Hindi-Urdu has two types of compound verbs: those that involve verbs in a series and those that involve a nominal and a verbal. Some examples of the former have already been given (see page 480), a few examples of the latter follow: *svikar kərna* 'acceptance do' or 'to accept', *pəsənd hona* 'liking be' or 'to like' (non-volitional), *pəsənd kərna* 'liking do' or 'to like' (volitional), *təng ana* 'torment come' or 'to be fed up'.

In the verbs-in-series type of compound verbs, usually the meaning of the whole is derived from the meaning of the first, or main, verb; the second, or explicator, verb performs the function of either restricting, or adding some specific shade of meaning to, the meaning of the main verb. Also, the explicator verb necessarily expresses the meaning 'a one-shot action or process'. For instance, *marna* can mean either 'hit' or 'kill', *mar ḍalna* 'hit/kill pour' means only 'kill'; *lɪkhna* means 'write', *lɪkh marna* 'write hit' means 'to dash off a few lines in a hurry/thoughtlessly'; *rəkhna* means 'keep, put', *rəkh čhoṛna'*, 'keep leave' means 'save'. The main explicator verbs are the following and they roughly signify the meanings described below:

ana 'come' occurs with intransitive verbs of motion and indicates that the action of the main verb is oriented towards a focal point which may be a person or which may be set in time or space; e.g. *vəh siṛhɪyā čəṛh ai* 'she came up the steps' and *vəh siṛhɪyõ se utər ai* 'she came down the steps'.

ǰana 'go' occurs with intransitive verbs of motion and other change-of-state verbs and indicates motion away from the focal point; with dative subject verbs, it indicates definitive meaning; and with transitive verbs, it indicates hurried, compulsive action; e.g. *vəh siṛhɪyā čəṛh gəi* 'she went up the steps', *raǰu ko kɪtab mɪl gəi* 'Raju got the book', *vəh gusse mẽ ǰane kya-kya lɪkh gəya* 'who knows what he dashed off in his anger!'

lena 'take' occurs with affective (see page 485) (transitive) verbs and indicates completive meaning; with other transitive verbs, it indicates a self-benefactive meaning; and with certain intransitive verbs, it indicates internal expression; e.g. *ʊsne kam kər lɪya* '(s)he completed (his/her) job', *mẽ ne thik soč lɪya hɛ* 'I have made a decision'.

dena 'give' occurs with transitive verbs other than affective verbs and indicates that the action is directed towards a beneficiary other than the agent of the action denoted by the main verb; and with intransitive verbs of expression, it indicates external expression; e.g. *ʊsne sara rəhəsy bəta dɪya* 'he divulged the whole secret', *sima zorõ se hãs di* 'Sima laughed loudly'.

ʊṭhna 'rise' occurs with intransitive and transitive verbs of punctual action and indicates suddenness; e.g. *vəh muǰhe dekhte hi ro ʊṭhi* 'she suddenly began to cry when she saw me'.

bɛṭhna 'sit' occurs with certain transitive verbs and indicates impudence; e.g. *vəh əpne 'bas' se ləṛ bɛtha* 'he fought with his boss'.

pəṛna 'fall' occurs with intransitive change-of-state verbs, and certain

verbs of expression, and indicates suddenness; e.g. *ləṛki bərf pər phɪsəl kər gɪr pəri* 'the girl slipped and fell on the ice'.

ḍalna 'pour' occurs with transitive verbs that express violent action and certain transitive verbs (*kər* 'do', *pərh* 'read', *lɪkh* 'write') and indicates violence; e.g. *ǰəldi se pətr lɪkh dalo!* 'write the letter quickly (get it over with)!'

rəkhna 'keep' occurs with certain transitive verbs and indicates a temporary state resulting from the action of the main verb; e.g. *mẽ ne khana pəka rəkha hɛ* 'I have cooked (and saved) the food'.

čhoṛna 'leave' occurs with certain transitive verbs and indicates dissociation of the agent with the result of the action; e.g. *pɪtaǰi ne meri pəṛhai ke lɪye pɛse rəkh čhoṛe hẽ* 'father has put aside money for my education'.

marna 'hit' occurs with very few verbs and indicates rash action; e.g. *kučh bhi lɪkh maro!* 'just write something!'

dhəməkna 'thump' occurs with *ana* 'come' and *ǰana* 'go' and indicates unwelcome arrival; e.g. *vəh sʊbəh-sʊbəh a dhəmka, mʊǰhe nəhane tək ka mɔka nəhi mɪla* 'he showed up very early, I did not even have time to shower'.

pəhʊ̃čna 'arrive' occurs with *ana* 'come' and *ǰana* 'go' and indicates arrival rather than motion; e.g. *šyam dɪlli ǰa pəhʊ̃ča* 'Shyam arrived in Delhi'.

nikəlna 'emerge' indicates sudden emergence from some enclosed space — real or imaginary; e.g. *ʊski ākhõ se āsu bəh nɪkle* 'tears began to flow from her eyes'.

4 Syntax

In this brief section on syntax, I will discuss mainly the verbal syntax of Hindi-Urdu after a few remarks on word order. The reason for this will become clearer as the discussion progresses.

Hindi-Urdu is a verb final language, i.e. the order of words in a sentence is subject, object and verb. Actually, the position of the verb is relatively more fixed than the position of any other constituent. Since most grammatical functions of nouns are indicated by the postpositions following them, the nominal constituents can be moved around freely for thematic purposes. The position of the verb is changed only in poetic or extremely affective style. Historically, word order was relatively free in Old Indo-Aryan, but became more fixed in Middle Indo-Aryan between AD 200 and 600.

In existential sentences, the locational/temporal adverbial comes first: *mez pər kɪtab hɛ* 'there is a book on the table', *kəl bəṛi ṭhəṇḍ thi* 'it was very cold yesterday'. The verb agrees with the unmarked noun in the sentence. In intransitive and non-perfective transitive sentences, where the subject is unmarked, the verb agrees with the subject, e.g. *ləṛke bɛṭhe* 'the boys sat', *ləṛki səmačar sʊn rəhi hɛ* 'the girl is listening (f.) to the news (m.)', *raǰu čay pita hoga* 'Raju (m.) must be drinking (m.) tea (f.)'. In transitive sentences in the perfective, where the subject is followed by the postposition *ne*, the verb does not agree with the subject. It agrees with the object if it is unmarked; if the object is followed by the postposition *ko*, the verb remains in its neutral form, i.e. third person singular masculine: cf. *raǰu ne kɪtab pəṛhi* 'Raǰu (m.) read (f.) the book (f.)', *əfsərõ ne əpni pətnɪyõ ko bʊlaya* 'the officers called (3rd sg. m.) their wives'. Not all transitive verbs require that their subjects be marked with the agentive postposition *ne*: e.g. *bolna* 'speak', *lana* 'bring' do not take *ne*, *səməǰhna* 'understand' can occur either with or without *ne*: *mɛ apki bat nəhĩ səmǰha* 'I do not understand you', *ap ne kya səmǰha?* 'what did you understand?' In the case of compound verbs, only if both the main and the explicator verbs require *ne* does the compound verb require *ne*: *šila ne dudh pɪya* 'Sheila drank the milk', *šila ne dudh lɪya* 'Sheila took the milk', *šila ne dudh pi lɪya* 'Sheila drank up the milk', but *šila dudh pi gəi* 'Sheila drank up the milk' since the intransitive verb *ǰa* 'go' is not a *ne* verb.

Semantically, Hindi-Urdu makes a distinction between volitional versus non-volitional verbs and affective versus non-affective verbs. A verb is volitional if it expresses an act that is performed by an actor/agent. A verb is affective if the act expressed by the verb is directed towards the actor/agent, i.e. it is self-benefactive. Ingestive verbs such as *khana* 'eat', *pina* 'drink' etc. are good examples of affective verbs in that it is the actor/agent of eating, drinking etc. who benefits from these acts. Verbs such as 'work', 'write' etc., on the other hand may be either self-benefactive or directed toward some other beneficiary. Typically, the explicator verb *lena* 'take' occurs with an affective verb, the explicator *dena* 'give' does not, i.e. sentences such as the following are ungrammatical in Hindi-Urdu: *ʊsne khana kha dɪya* 'he/she ate for someone else' because *khana* 'eat' is an ingestive verb whereas the explicator *dena* 'give' indicates that the beneficiary is someone other than the actor/agent of the main verb. Verbs such as *girna* 'fall', *ǰana* 'go' etc. express self-directed actions, hence are affective.

These distinctions are important for the verbal syntax of Hindi-Urdu. Transitivity, volitionality and affectiveness do not necessarily coincide. For instance, *sona* 'sleep' is intransitive, volitional and affective, *sikhna* 'learn' is transitive, volitional and affective, *gɪrna* 'fall' is intransitive, non-volitional and affective, *ǰana* 'go' is intransitive, volitional and affective. Only the

affective verbs participate in the compound verbal construction with *lena* 'take' as the explicator, only volitional verbs occur in the passive construction (Kachru 1980; 1981).

In many cases, verbs in Hindi-Urdu come in related forms so that the stative versus active and volitional versus non-volitional meanings can be expressed by varying the syntactic constructions. For instance, the verb *mɪlna* can mean both 'to run into someone' (accidental meeting) or 'to go see someone' (deliberate meeting). In the first case, the verb is used with a dative subject and the object of meeting is unmarked, in the second case, the subject is unmarked and the object is marked with a comitative postposition *se*, e.g. *kəl bazar ǰate hʊe mʊǰhe ram mɪla tha* 'yesterday while going to the market I ran into Ram', *kəl mẽ ram se ʊske dəftər mẽ mɪla tha* 'yesterday I met Ram in his office'. In a large number of cases, the intransitive verb denotes non-volitional action and if the actor is to be expressed, it is expressed with the instrumental postposition *se*, e.g. *apka šiša mʊǰhse ṭuṭ gəya* 'your mirror got broken by me'. The deliberate action is expressed with the related transitive verb in the agentive construction, e.g. *ɪs šərarti bəčče ne apka šiša toṛ ḍala* 'this naughty child broke your mirror'. Most intransitive and all dative subject verbs are either stative or change-of-state verbs and are non-volitional. Hindi-Urdu has sets of stative, change-of-state and active verbs of the following types:

Stative	*Change-of-state*	*Active*
khʊla hona 'be open'	khʊlna	kholna
krʊddh hona 'be angry'	krodh ana	krodh kərna
yad hona 'remember'	yad ana	yad kərna
pəsənd hona 'like'	pəsənd ana	pəsənd kərna

Note that the stative verbs are usually made up of an adjective or past participle and the verb 'be', the change-of-state verbs are either lexical verbs or compounds made up of a nominal and the verb 'become' or 'come', and the active is either a causal verb morphologically derived from the intransitive or a compound made up of a nominal and the verb 'do' (or a small set of other active transitive verbs).

This, however, does not mean that all intransitive verbs in Hindi are of the above types. There are active intransitive verbs such as the verbs of motion (*ǰa* 'go', *čəl* 'move' etc.), verbs of expression (*hãs* 'laugh', *ro* 'cry' etc.) and others. Note that verbal compounding is also exploited to reduce volitionality of verbs, e.g. *ro pəṛna* 'cry + fall = to burst out crying', *bol ʊṭhna* 'speak + rise = to blurt out' etc.

The non-volitional intransitive sentence above (*apka šiša mʊǰhse ṭuṭ gəya* 'your mirror got broken by me') has been translated into English with the passive; it is, however, not a passive construction in Hindi-Urdu. The passive in Hindi-Urdu is formed by marking the agent of the active sentence, if retained, with the instrumental postposition *se* and using the perfect stem of the verb and the auxiliary *ǰa* 'go' which takes all the tense-aspect endings: e.g. *ram ne khana nəhī khaya* 'Ram did not eat' vs. *ram se khana nəhī khaya gəya* 'Ram was not able to eat'. The translation equivalent of the Hindi-Urdu passive in English points to an interesting fact about this construction. If the agent is retained and marked with the instrumental postposition, the passive sentence is usually interpreted as a statement about the capability of the agent; if, however, the agent is deleted, the passive sentence has a meaning similar to that of English. That is, the sentence is interpreted as being about the object in the active sentence and the agent is either unknown or not important enough to be mentioned (Guru 1920; Kachru 1980).

In addition to the present and past participles, there are two other participles in Hindi which are used a great deal: the conjunctive participle which is formed by adding the form *kər* to the root of the verb and the agentive participle which is formed by adding the suffix *-vala* to the oblique form of the verbal noun, e.g. *lɪkhnevala* 'writer', *ǰanevala* 'one who goes', *sonevala* 'one who sleeps', *ʊgnevala* 'that which rises or grows', etc. This suffix has become a part of the English lexicon in the form *wallah* and is used extensively in Indian English and the native varieties of English, especially in the context of topics related to India. Forms such as *Congresswallah* ('one belonging to the Indian National Congress'), *Bombaywallah* ('one from Bombay') are common in literature dealing with India.

The syntax of Hindi-Urdu differs from that of English most noticeably in the use of the participles. For instance, the preferred constructions for modifying nouns or conjoining clauses are the participles: the present, past and agentive for modifying nouns and the conjunctive participle for conjoining clauses. Compare the following Hindi sentences with their English translations: *vəh gẽd khelte hʊe bəččõ ko dekh rəha tha* 'he was observing the children (who were) playing ball'; *tʊmhẽ mohən ki lɪkhi hʊi kəvɪtaẽ pəsənd hẽ?* 'do you like the poems written by Mohan?'; *mʊǰhe bat bat pər ronewale bəčče bɪlkʊl pəsənd nəhī* 'I do not like children who cry at every thing'; *vəh ghər a kər so gə ya* 'he came home and went to sleep'. Both the present and the past participles are used adjectivally as well as adverbially, cf. *mã ne rote hʊe bəčče ko god mẽ ʊṭha lɪya* 'Mother picked up the child who was crying' vs. *vəh rote hʊe bhag gəya* 'he ran away, crying' and *mẽ vəhã bɛṭhi hʊ lərki ko nəhī ǰanti* 'I don't know the girl seated over there' vs. *lərki vəhã bɛṭhi (hui) pətr lɪkh rəhi hɛ* 'the girl is writing a letter sitting there'. The agentive participle is used both as an agentive noun, e.g. *(gaṛi) cəlanevala* 'driver (of a vehicle)' and as an adjective, e.g. *bharət se anevale čhatr* 'the students who come from India'. The conjunctive participle is used to express the meanings of sequential action, related action, cause-effect relationship and purpose adverbial, e.g. *vəh hɪndi pəṛh kər khelne ǰaega* 'he will go to play after studying Hindi', *vəh kud kər upər a gəi* 'she jumped and came up', *həm ne ʊse pɛse de kər xuš kər lɪya* 'we

pleased him by giving him money', *ǰəldi se bazar ǰakər dudh le ao* 'go quickly to the market and bring some milk' (Kachru 1980).

Although the participial constructions are preferred in Hindi-Urdu, there are linguistically determined environments where full relative and other types of subordinate and conjoined clauses are used. The relative clause, unlike in English, is not a constituent of the noun phrase. It may either precede or follow the main clause as in the following: *ǰo lərka vəhã bɛṭha hɛ vəh mera bhai hɛ* or *vəh lərka mera bhai hɛ ǰo vəhã bɛṭha hɛ* 'the boy who is seated there is my brother'. Note that, depending upon the order of the relative and the main clause, either the noun in the subordinate or the main clause is deleted, i.e. the above are the results of deleting the noun in parentheses in the following: *ǰo lərka vəhã bɛṭha hɛ vəh (lərka) mera bhai hɛ* or *vəh lərka mera bhai hɛ ǰo (lərka) vəhã bɛṭha hɛ*. The relative marker *ǰo* (obl. sg. *ǰɪs*, obl. pl. *ǰɪn*, special forms with *ne* and *ko*, *ǰɪnhõne* and *ǰɪnhẽ*) and the correlative marker *vəh*, which is identical to the remote demonstrative/third person pronoun, function like a determiner to their respective head nouns. Both the head nouns may be retained in the case of an emphatic construction; in normal speech/writing, however, the second instance is deleted. Under the influence of Persian and later, English, the relative clause is sometimes positioned following the head noun, e.g. *vəh lərka ǰo vəhã bɛṭha hɛ mera bhai hɛ*; in this case, the second instance of the noun (following *ǰo*) must be deleted.

Earlier, it has been said that the nominal constituents of a sentence in Hindi-Urdu can be moved around freely for thematic purposes. Usually, the initial element in a sentence in Hindi coincides with the theme. The focus position in Hindi is identified with the position just before the main verb. In addition to manipulating the word order, heavy sentence stress and certain particles are used to indicate focus, e.g. *'ram' ne mohən ko piṭa* 'it was Ram who hit Mohan', *šila hi ne yəh bat kəhi thi* 'it was Sheila who had said this', *sima to čəli gəi*, 'as for Sima, she has left', where the item in quotes in the first sentence and the items followed by the particles *hi* and *to* in the second and the third sentence respectively are under focus. As the initial position is not the favoured device for indicating focus, the interrogative pronouns in Hindi-Urdu do not necessarily occur sentence-initially; compare the Hindi-Urdu sentences with their English equivalents, *ap kya pərh rəhe hẽ?'* 'what are you reading?', *vəh kəl kəhã gəya tha?* 'where did he go yesterday?', *ɪn mẽ se ap ko kɔn si kɪtab pəsənd hɛ?* 'which of these books do you like?'.

To sum up, Hindi-Urdu differs from its European cousins typologically in several respects. Phonologically, aspiration, retroflexion, nasal vowels and lack of distinctive stress mark Hindi-Urdu as very different from English. Morphologically, the gender and case distinctions and the devices of reduplication and echo-compounding exemplify the major differences between the two languages. Syntactically, the word order differences are striking. So is the fact that Hindi-Urdu makes certain semantic distinctions which are not made as clearly in English, viz. volitionality and affectiveness. These distinctions result in a closer correspondence between semantic and syntactic grammatical roles that nominal constituents have in a sentence, e.g. all agentive (*-ne*-marked) subjects are agents, all dative (*ko*-marked) subjects are experiencers, and so on. Many of these characteristics of Hindi-Urdu are shared by not only the other Indo-Aryan but also the Dravidian and other languages of India.

Bibliography

The standard reference grammar for modern standard Hindi is Guru (1920). Other reference grammars are Sharma (1958) and McGregor (1972), the latter directed to the needs of learners of Hindi. Kellogg (1875) describes Hindi and the major dialects of the Hindi area, and contains a good introduction to Hindi prosody; data are drawn mostly from literary texts of the period and the work is hence dated. For Urdu, see Bailey (1956). For an account of the parallel development of Hindi-Urdu see Rai (1984).

Ohala (1983) is a phonological description of Hindi, while Kachru (1980) describes syntactic constructions of Hindi in non-technical language. Verma (1933) is a brief sketch of the history of the Hindi language. Kachru (1981) contains a supplement on transplanted varieties of Hindi-Urdu and one on transitivity in Hindi-Urdu. Bhatia (1987) discusses the native and non-native grammatical tradition.

References

Bailey, T.G. 1956. *Teach Yourself Urdu* (English Universities Press, London)

Bhatia, T.K. 1987. *A History of Hindi (Hindustani) Grammatical Tradition* (E.J. Brill, Leiden)

Guru, K.P. 1920. *Hindi vyākaraṇ* (Kashi Nagri Pracharini Sabha, Banaras)

Kachru, Y. 1980. *Aspects of Hindi Grammar* (Manohar Publications, New Delhi)

—— 1981. 'Dimensions of South Asian Linguistics', *Studies in the Linguistic Sciences*, vol. 11, no. 2

Kellogg, S.H. 1875. *A Grammar of the Hindi Language* (Routledge and Kegan Paul, London)

McGregor, R.S. 1972. *Outline of Hindi Grammar* (Oxford University Press, London)

Ohala, M. 1983. *Aspects of Hindi Phonology* (Motilal Banarsidass, Delhi)

Rai, A. 1984. *A House Divided* (Oxford University Press, Delhi)

Sharma, A. 1958. *A Basic Grammar of Modern Hindi* (Government of India, Ministry of Education and Scientific Research, Delhi)

Verma, Dh. 1933. *Hindi bhaṣa ka itihas* (Hindustani Academy, Allahabad)

23 Bengali

M. H. Klaiman

1 Historical and Genetic Setting

Bengali, together with Assamese and Oriya, belongs to the eastern group within the Magadhan subfamily of Indo-Aryan. In reconstructing the development of Indo-Aryan, scholars hypothetically posit a common parent language from which the modern Magadhan languages are said to have sprung. The unattested parent of the Magadhan languages is designated as Eastern or Magadhi Apabhraṁśa, and is assigned to Middle Indo-Aryan. Apart from the eastern languages, other modern representatives of the Magadhan subfamily are Magahi, Maithili and Bhojpuri.

Within the eastern group of Magadhan languages, the closest relative of Bengali is Assamese. The two share not only many coincidences of form and structure, but also have in common one system of written expression, on which more details will be given later.

Historically, the entire Magadhan group is distinguished from the remaining Indo-Aryan languages by a sound change involving sibilant coalescence. Specifically, there occurred in Magadhan a falling together of three sibilant elements inherited from common Indo-Aryan, dental /s/, palatal /š/ and retroflex /ṣ/. Among modern Magadhan languages, the coalescence of these three sounds is manifested in different ways; e.g. the modern Assamese reflex is the velar fricative /x/, as contrasted with the palatal /š/ of Modern Bengali.

The majority of Magadhan languages also show evidence of historical regression in the articulation of what was a central vowel /ă/ in common Indo-Aryan; the Modern Bengali reflex is /ɔ/.

Although the Magadhan subfamily is defined through a commonality of sound shifts separating it from the rest of Indo-Aryan, the three eastern languages of the subfamily share one phonological peculiarity distinguishing them from all other modern Indo-Aryan languages, both Magadhan and non-Magadhan. This feature is due to a historical coalescence of the long and short variants of the high vowels, which were distinguished in common Indo-Aryan. As a result, the vowel inventories of Modern Bengali, Assamese and Oriya show no phonemic distinction of /ĭ/ and /ī/, /ŭ/ and /ū/.

Moreover, Assamese and Bengali are distinguished from Oriya by the innovation of a high/low distinction in the mid vowels. Thus Bengali has /æ/ as well as /e/, and /ɔ/ as well as /o/. Bengali differs phonologically from Assamese principally in that the latter lacks a retroflex consonant series, a fact which distinguishes Assamese not just from Bengali, but from the majority of modern Indo-Aryan languages.

Besides various phonological characteristics, there are certain grammatical features peculiar to Bengali and the other Magadhan languages. The most noteworthy of these features is the absence of gender, a grammatical category found in most other modern Indo-Aryan languages. Bengali and its close relative Assamese also lack number as a verbal category. More will be said on these topics in the section on morphology, below.

Writing and literature have played no small role in the evolution of Bengali linguistic identity. A common script was in use throughout eastern India centuries before the emergence of the separate Magadhan vernaculars. The Oriya version of this script underwent special development in the medieval period, while the characters of the Bengali and Assamese scripts coincide with but a couple of exceptions.

Undoubtedly the availability of a written form of expression was essential to the development of the rich literary traditions associated not just with Bengali, but also with other Magadhan languages such as Maithili. However, even after the separation of the modern Magadhan languages from one another, literary composition in eastern India seems to have reflected a common milieu scarcely compromised by linguistic boundaries. Although vernacular literature appears in eastern India by AD 1200, vernacular writings for several centuries thereafter tend to be perceived as the common inheritance of the whole eastern area, more so than as the output of individual languages.

This is clearly evident, for instance, in the case of the celebrated Buddhist hymns called the *Caryāpada*, composed in eastern India roughly between AD 1000 and 1200. Though the language of these hymns is Old Bengali, there are reference works on Assamese, Oriya and even Maithili that treat the same hymns as the earliest specimens of each of these languages and their literatures.

Bengali linguistic identity is not wholly a function of the language's genetic affiliation in the Indo-Aryan family. Eastern India was subjected to Aryanisation before the onset of the Christian era, and therefore well before the evolution of Bengali and the other Magadhan languages. Certain events of the medieval era have had a greater significance than Aryanisation in the shaping of Bengali linguistic identity, since they furnished the prerequisites of Bengali regional and national identity.

Among these events, one of the most crucial was the establishment of Islamic rule in the early thirteenth century. Islamisation led to six hundred

years of political unity in Bengal, under which it was possible for a distinctly national style of literary and cultural expression to evolve, more or less unaffected by religious distinctions. To be sure, much if not all early popular literature in Bengali had a sacred basis; the early compositions were largely translations and reworkings of Hindu legends, like the Krishna myth cycle and the *Rāmāyaṇa* religious epic. However, this material seems to have always been looked upon more as a product of local than of sectarian tradition. From the outset of their rule, the Muslim aristocracy did little to discourage the composition of literature on such popular themes; on the contrary, they often lent their patronage to the authors of these works, who were both Muslim and Hindu. Further, when in the sixteenth and seventeenth centuries Islamic writers ultimately did set about creating a body of sectarian, didactic vernacular literature in Bengali, they readily adapted the originally Hindu motifs, themes and stories that had become part of the local cultural tradition.

The relative weakness of religious identity in Bengali cultural institutions is perhaps best interpreted in light of a major event which occurred concomitant to the rise of Islamic rule. This event was a massive shift in the course of the Ganges River between the twelfth and sixteenth centuries AD. Whereas it had earlier emptied into the Bay of Bengal nearly due south of the site of present-day Calcutta, the river gradually approached and eventually became linked with the Padma River system in the territory today called Bangladesh. The shift in the Ganges has been one of the greatest influences upon material history and human geography in eastern India; for, prior to the completion of the river's change of course, the inhabitants of the eastern tracts had been virtually untouched by civilisation and sociocultural influences from without, whether Islamic or Hindu. Over the past four centuries, it is the descendants of the same people who have come to make up the majority of speakers of the Bengali language; so that the basis of their Bengali identity is not genetic and not religious, but linguistic. That the bulk of the population perceives commonality of language as the principal basis of its social unity is clear from the name taken by the new nation-state of eastern Bengal following the 1971 war of liberation. In the proper noun *Bangladesh* (composed of *bāṅglā* plus *deśa*, the latter meaning 'country'), the first part of the compound does not mean the Bengali people or the territory of Bengal; the term *bāṅglā* specifically refers, rather, to the Bengali language.

The Muslim aristocracy that ruled Bengal for some six centuries was supplanted in the eighteenth century by new invaders, the British. Since the latter's withdrawal from the subcontinent in 1947, the community which identifies itself as Bengali has been divided between two sovereign political entities. However, the Bengali language continues to be spoken throughout Bengal's traditional domains, and on both sides of the newly-imposed international boundary. Today, Bengali is one of the official regional

speeches of the Indian Union, a status which is also enjoyed by the other eastern Magadhan languages, Oriya and Assamese. Among the three languages, the one which is currently in the strongest position is Bengali, since it alone also has the status of a national language outside India's present borders. In India, about eight per cent of the overall population, or some 55 million people per 1981 census figures, speak Bengali. The great bulk of these speakers reside in West Bengal, the Indian state contiguous to Bangladesh. At the same time, in Bangladesh, 1980 census figures report a population of nearly ninety million, of whom over 95 per cent are Bengali speakers. Thus the combined community of Bengali speakers in India and Bangladesh approaches 145 million, a larger body of native speakers than currently exists for French.

2 Orthography and Sound System

The writing system of Modern Bengali is derived from Brāhmī, an ancient Indian syllabary. Brāhmī is also the source of all the other native Indian scripts (including those of the modern South Indian languages) as well as of Devanāgarī, a script associated with classical Sanskrit and with a number of the modern Indo-Aryan languages.

The scripts of the modern eastern Magadhan languages (Oriya, Assamese and Bengali) are based on a system of characters historically related to, but distinct from, Devanāgarī. The Bengali script is identical to that of Assamese except for two characters; while the Oriya script, though closely related historically to the Bengali-Assamese script, is quite distinctive in its appearance.

Like all Brāhmī-derived scripts, Bengali orthography reads from left to right, and is organised according to syllabic rather than segmental units.

Table 23.1: Bengali Script

Vowel Segments *Special name of character, if any*	*Independent form*	*Combining form (shown with the sign kɔ)*	*Transliteration*
	অ	ক	ɔ
	আ	কা	a
hrɔsso i	ই	কি	i
dirgho i	ঈ	কী	ī
hrɔsso u	উ	কু	u
dirgho u	ঊ	কূ	ū
ri	ঋ	কৃ	ri
	এ	কে	e
	ঐ	কৈ	oy
	ও	কো	o
	ঔ	কৌ	ow

Consonant Segments	*Ordinary form*	*Special form(s)*	*Transliteration (so-called 'inherent vowel' not represented)*
	ক		k
	খ		kh
	গ		g
	ঘ		gh
	ঙ	ং	ṅ
	চ		c
	ছ		ch
	জ		j
	ঝ		jh
	ঞ		ñ
	ট		ṭ
	ঠ		ṭh
	ড		ḍ
	ড়		ṛ
	ঢ		ḍh
	ঢ়		ṛh
	ণ		ṇ
	ত	ৎ	t
	থ		th
	দ		d
	ধ		dh
	ন		n
	প		p
	ফ		ph
	ব		b
	ভ		bh
	ম		m
ɔntostho jɔ	য		j
ɔntostho ɔ	য়	◌	y, w
	র		r
	ল		l
talobbo sɔ	শ		ś
murdhonno sɔ	ষ		ṣ
donto sɔ	স		s
	হ	ঃ	h
Special diacritics			
cɔndrobindu	ঁ		◌̃
hɔsonto	্		◌

Accordingly, a special diacritic or character is employed to represent a single consonant segment in isolation from any following vowel, or a single vowel in isolation from any preceding consonant. Furthermore, the writing system of Bengali, like Devanāgarī, represents characters as hanging from a superimposed horizontal line and has no distinction of upper and lower cases.

Table 23.1 sets out the Bengali script according to the traditional ordering of characters, with two special diacritics listed at the end. Most Bengali characters are designated according to the pronunciation of their independent or ordinary form. Thus the first vowel character is called *ɔ*, while the first consonant character is called *kɔ*. The designation of the latter is such, because the corresponding sign in isolation is read not as a single segment, but as a syllable terminating in /ɔ/, the so-called 'inherent vowel'. Several Bengali characters are not designated by the pronunciation of their independent or ordinary forms; their special names are listed in the leftmost column of table 23.1. Among the terms used in the special designations of vowel characters, *hrɔsso* literally means 'short' and *dirgho* 'long'. Among the terms used in the special designations of consonant characters, *talobbo* literally means 'palatal', *murdhonno* 'retroflex', and *donto* 'dental'. These terms are used, for historical reasons, to distinguish the names for the three sibilant characters. The three characters (transliterated *ś*, *ṣ* and *s*) are used to represent a single non-obstruent sibilant phoneme in Modern Bengali. This phoneme is a palatal with a conditioned dental allophone; further discussion will be given below. It might be pointed out that another Bengali phoneme, the dental nasal /n/, is likewise represented in orthography by three different characters, which are transliterated *ñ*, *ṇ*, and *n*.

In Bengali orthography, a vowel sign normally occurs in its independent form only when it is the first segment of a syllable. Otherwise, the combining form of the vowel sign is written together with the ordinary form of a consonant character, as illustrated in table 23.1 for the character *kɔ*. There are a few exceptional cases: for instance, the character *hɔ* when written with the combining form of the sign *ri* appears not as হৃ , but as হৃ (pronounced [hri]). The character *rɔ* combined with *dirgho u* is written not as র্ , but as রূ [ru]. The combination of *talobbo sɔ* with *hrɔsso u* is optionally represented either as শু or as শু (both are pronounced [šu]), while *gɔ*, *rɔ* and *hɔ* in combination with *hrɔsso u* yield the respective representations গু [gu], রু [ru], and হু [hu].

Several of the consonant characters in Bengali have special forms designated in table 23.1; their distribution is as follows. The characters *ṅɔ* and *tɔ* occur in their special forms when the consonants they represent are the final segments of phonological syllables. Thus /baṅla/ 'Bengali language' is written বাংলা , while /štɔt/ 'true' is written সৎ .

The character *ɔntostho* ɔ has a special form listed in table 23.1; the name of this special form is *jɔ phɔla*. Generally, *jɔ phɔla* is the form in which *ɔntostho* ɔ occurs when combined with a preceding ordinary consonant sign, as in ত্যাগ [tæg] 'renunciation'. When combined with an ordinary consonant sign in non-initial syllables, *jɔ phɔla* tends to be realised as gemination of the consonant segment, as in গ্রাম্য [grammo] 'rural'. The sign *ɔntostho* ɔ in its ordinary form is usually represented intervocalically, and generally realised phonetically as a front or back high or mid semi-vowel. Incidentally, the

character *ɔntostho ɔ* in its ordinary form is not to be confused with the similar looking character that precedes it in table 23.1, the *ɔntostho jɔ* character. This character has the same phonemic realisation as the consonant sign *jɔ* (listed much earlier in table 23.1), and is transliterated in the same way. While *jɔ* and *ɔntostho jɔ* have the same phonemic realisation, they have separate historical sources; and the sign *ɔntostho jɔ* occurs today in the spelling of a limited number of Bengali lexemes, largely direct borrowings from Sanskrit.

The sign *rɔ* exhibits one of two special forms when written in combination with an ordinary consonant sign. In cases where the ordinary consonant sign represents a segment which is pronounced before /r/, then *rɔ* appears in the combining form *rɔ phɔla*; to illustrate: প্রেত [pret] 'ghost, evil spirit'. In cases where the sound represented by the ordinary consonant sign is realised after /r/, *rɔ* appears in the second of its combining forms, which is called *reph*; as in অর্থ [ɔrtho] 'value'.

The sign *hɔ* has a special form, listed in table 23.1, which is written word-finally or before a succeeding consonant in the same syllable. In neither case, however, is the special form of *hɔ* very commonly observed in Bengali writing.

Two special diacritics are listed at the end of table 23.1. The first of these, *cɔndrobindu*, represents the supersegmental for nasalisation, and is written over the ordinary or combining form of any vowel character. The other special diacritic, called *hɔsonto*, is used to represent two ordinary consonant signs as being realised one after another, without an intervening syllabic, in the same phonological syllable; or to show that an ordinary consonant sign written in isolation is to be realised phonologically without the customary 'inherent vowel'. Thus: বাক্ [bak] 'speech', বাক্‌শক্তি [bakšokti] 'power of speech'. In practice, the use of this diacritic is uncommon, except where spelling is offered as a guide to pronunciation; or where the spelling of a word takes account of internal morpheme boundaries, as in the last example.

Table 23.1 does not show the representation of consonant clusters in Bengali orthography. Bengali has about two dozen or so special *soṅjukto* (literally 'conjunct') characters, used to designate the combination of two, or sometimes three, ordinary consonant signs. In learning to write Bengali, a person must learn the *soṅjukto* signs more or less by rote.

Before considering the sound system of Bengali, it should be mentioned that the spelling of Bengali words is well standardised, though not in all cases a strict guide to pronunciation. There are two especially common areas of inconsistency. One involves the representation of the sound [æ]. Compare the phonetic realisations of the following words with their spellings and transliterations: [æto] এত (transliterated *etɔ*) 'so much, so many'; [bæsto] ব্যস্ত (transliterated *byɔstɔ*) 'busy'; and [læj] ল্যাজ (transliterated *lyajɔ*) 'tail'. The sound [æ] can be orthographically represented in any of the three ways illustrated, and the precise spelling of any word containing this sound must accordingly be memorised.

Another area of inconsistency involves the realisation of the 'inherent vowel'. Since, as mentioned above, the diacritic *hɔsonto* (used to indicate the absence of the inherent vowel) is rarely used in practice, it is not always clear whether an unmodified ordinary consonant character is to be read with or without the inherent vowel. Compare, for example, [kɔto] কত (transliterated *kɔtɔ*) 'how much/how many' with [mɔt] মত (transliterated *mɔtɔ*) 'opinion'. This example makes it especially clear that Bengali spelling is not an infallible guide to pronunciation.

The segmental phonemes (oral vowels and consonants) of the standard dialect of Bengali are set forth in table 23.2. As table 23.2 makes clear, the feature of aspiration is significant for obstruents and defines two phonemically distinct series, the unaspirates and the aspirates. Though not represented in the table since it is non-segmental, the feature of nasalisation is nonetheless significant for vowels and similarly defines two phonemically distinct series. Thus in addition to the oral vowels as listed in table 23.2, Bengali has the corresponding nasalised vowel phonemes /ɔ̃/, /ã/, /æ̃/, /õ/, /ẽ/, /ũ/ and /ĩ/.

Table 23.2: Segmental Phonemes of Bengali

Consonants

	Labial	*Dental*	*Retroflex*	*Palatal*	*Velar*	*Post-velar*
Obstruents						
voiceless:						
unaspirated	p	t	ṭ	c	k	
aspirated	ph	th	ṭh	ch	kh	
voiced:						
unaspirated	b	d	ḍ	j	g	
aspirated	bh	dh	ḍh	jh	gh	
Nasals	m	n			ṅ	
Flaps		r	ṛ			
Lateral		l				
Spirants				s		h

Vowels

	Front	*Back*
High	i	u
High mid	e	o
Low mid	æ	ɔ
Low	a	

The phonemic inventory of modern standard Bengali marks it as a fairly typical Indo-Aryan language. The organisation of the consonant system in terms of five basic points of articulation (velar, palatal, retroflex, dental and labial) is characteristic, as is the stop/flap distinction in the retroflex series.

(Hindi-Urdu, for instance, likewise has several retroflex stop phonemes and retroflex flaps.) Also typically Indo-Aryan is the distinctive character of voicing in the Bengali obstruent inventory, along with the distinctive character of aspiration. The latter feature tends, however, to be suppressed preconsonantally, especially in rapid speech. Moreover, the voiced labial aspirate /bh/ tends to be unstable in the pronunciation of many Bengali speakers, often approximating to a voiced labial continuant [v].

In the consonant inventory, Bengali can be regarded as unusual only in having a palatal sibilant phoneme in the absence of a dental sibilant. The historical background of this has been discussed in the preceding section. The phoneme in question is realised as a palatal [š] in all environments, except before the segments /t/, /th/, /n/, /r/, and /l/, where it is realised as a dental, i.e. as [s]. For simplicity, this Bengali sibilant is represented as *s* in the remainder of this chapter.

Nasalisation as a distinctive non-segmental feature of the vowel system is typical not only of Bengali but of modern Indo-Aryan languages generally. In actual articulation, the nasality of the Bengali nasalised vowel segments tends to be fairly weak, and is certainly not as strong as the nasality of vowels in standard French.

The most interesting Modern Bengali phonological processes involve the vowel segments to the relative exclusion of the consonants. One process, Vowel Raising, produces a neutralisation of the high/low distinction in the mid vowels, generally in unstressed syllables. Given the stress pattern of the present standard dialect, which will be discussed later, Vowel Raising generally applies in non-word-initial syllables. Evidence for the process is found in the following alternations:

məl	'dirt'	ɔmol	'pure'
sɔ	'hundred'	ækso	'one hundred'
æk	'one'	ɔnek	'many'

A second phonological process affecting vowel height is very significant because of its relationship to morphophonemic alternations in the Bengali verbal base. This process may be called Vowel Height Assimilation, since it involves the assimilation of a non-high vowel (other than /a/) to the nearest succeeding vowel segment within the phonological word, provided the latter has the specification [+high]. Outside the area of verbal morphophonemics, the evidence for this process principally comes from the neutralisation of the high/low distinction in the mid vowels before /i/ or /u/ in a following contiguous syllable. Some alternations which illustrate this process are:

æk	'one'	ekṭi	'one' (plus classifier -ṭi)
lɔjja	'shame'	lojjito	'ashamed'
nɔṭ	'actor'	noṭi	'actress'
æk	'one'	ekṭu	'a little, a bit'
tɔbe	'then'	tobu	'but (then)'

At this point it will be useful to qualify the observation drawn earlier that Bengali is — phonologically speaking — a fairly typical Indo-Aryan language. It is true that most of the segments in the Modern Bengali sound system can be traced more or less directly to Old Indo-Aryan. However, the retroflex flap /ṛ/ of the former has no counterpart in the latter, and its presence in modern standard Bengali (and in some of its sisters) is due to a phonological innovation of Middle Indo-Aryan. Furthermore, while the other retroflex segments of Modern Bengali (/ṭ/, /ṭh/, /ḍ/, /ḍh/) have counterparts in the Old Indo-Aryan sound system, their overall frequency (phonetic load) in Old Indo-Aryan was low. On the other hand, among the modern Indo-Aryan languages, it is Bengali (along with the other Magadhan languages, especially the eastern Magadhan languages) which demonstrates a comparatively high frequency of retroflex sounds. Some external, i.e. non-Aryan influence on the diachronic development of the Bengali sound system is suggested. Such a hypothesis ought logically to be tied in with the observation in the earlier section of this essay that the numerical majority of Bengali speakers represents what were, until recent centuries, culturally unassimilated tribals of eastern Bengal, about whose prior linguistic and social history not much is known.

Further evidence of probable non-Aryan influence in the phonology is to be found in the peculiar word stress pattern of Modern Bengali. Accent was phonemic only in very early Old Indo-Aryan, i.e. Vedic (see page 456). Subsequently, however, predictable word stress has typified the Indo-Aryan languages; the characteristic pattern, moreover, has been for the stress to fall so many morae from the end of the phonological word. Bengali word stress, though, is exceptional. It is non-phonemic and, in the standard dialect, there is a strong tendency for it to be associated with word-*initial* syllables. This pattern evidently became dominant after AD 1400, or well after Bengali acquired a linguistic identity separate from that of its Indo-Aryan sisters. What this and other evidence may imply about the place of Bengali within the general South Asian language area is an issue to be further pursued toward the end of this essay.

3 Morphology

Morphology in Modern Bengali is non-existent for adjectives, minimal for nouns and very productive for verbs. Loss or reduction of the earlier Indo-Aryan adjective declensional parameters (gender, case, number) is fairly typical of the modern Indo-Aryan languages; hence the absence of adjectival morphology in Modern Bengali is not surprising. Bengali differs from many of its sisters, however, in lacking certain characteristic nominal categories. The early Indo-Aryan category of gender persists in most of the modern languages, with the richest (three-gender) systems still to be found in some of the western languages, such as Marathi. Early stages of the

Magadhan languages (e.g. Oriya, Assamese and Bengali) also show evidence of a gender system. However, the category is no longer productive in any of the modern Magadhan languages. In Modern Bengali, it is only in a few relic alternations (e.g. the earlier cited pair *nɔṭ* 'actor'/*noṭi* 'actress') that one observes any evidence today for the system of nominal gender which once existed in the language.

The early Indo-Aryan system of three number categories has been reduced in Modern Bengali to a singular/plural distinction which is marked on nouns and pronouns. The elaborate case system of early Indo-Aryan has also been reduced in Modern Bengali as it has in most modern Indo-Aryan languages. Table 23.3 summarises the standard Bengali declension for full nouns (pronouns are not given). Pertinent parameters not, however, revealed in this table are animacy, definiteness and determinacy. Generally, the plural markers are added only to count nouns having animate or definite referents; otherwise plurality tends to be unmarked. Compare, e.g. *jutogulo dɔrkar* 'the (specified) shoes are necessary' versus *juto dɔrkar* '(unspecified) shoes are necessary'. Further, among the plurality markers listed in table 23.3, *-gulo* (nominative), *-guloke* (objective), *-gulor* (genitive) and *-gulote* (locative-instrumental) are applicable to nouns with both animate and inanimate referents, while the other markers cooccur only with animate nouns. Hence: *chelera* '(the) boys', *chelegulo* '(the) boys', *jutogulo* 'the shoes', but **jutora* 'the shoes'.

Table 23.3: Bengali Nominal Declension

	Singular	*Plural*
Nominative	Ø	-ra/-era; -gulo
Objective	-ke	-der(ke)/-eder(ke); -guloke
Genitive	-r/-er	-der/-eder; -gulor
Locative-Instrumental	-te/-e *or* -ete	-gulote

The Bengali case markers in table 23.3 which show an alternation of form (e.g. *-r/-er*, *-te/-e* or *-ete*, *-der(ke)/-eder(ke)*, etc.) are phonologically conditioned according to whether the forms to which they are appended terminate in a syllabic or non-syllabic segment respectively. Both *-eder(ke)* and *-ete* are, however, currently rare. The usage of the objective singular marker *-ke*, listed in table 23.3, tends to be confined to inanimate noun phrases having definite referents and to definite or determinate animate noun phrases. Thus compare *kichu (*kichuke) caichen* 'do you want something?' with *kauke (*kau) caichen* 'do you want someone?'; but: *pulis caichen* 'are you seeking a policeman/some policemen?' versus *puliske caichen* 'are you seeking the police?'.

Bengali subject-predicate agreement will be covered in the following section on syntax. It bears mentioning at present, however, that the sole

parameters for subject-verb agreement in Modern Bengali are person (three are distinguished) and status. Inflectionally, the Bengali verb is marked for three status categories (despective/ordinary/honorific) in the second person and two categories (ordinary/honorific) in the third. It is notable that the shapes of the honorific inflectional endings are modelled on earlier Indo-Aryan plural inflectional markers. Table 23.4 lists the verbal inflection of modern standard Bengali.

Table 23.4: Bengali Verbal Inflection

	1st person	*2nd person despective*	*2nd person ordinary*	*3rd person ordinary*	*Honorific (2nd, 3rd persons)*
Present imperative	–	Ø	-o	-uk	-un
Unmarked indicative and *-(c)ch-* stems	-i	-is	-o	-e	-en
-b- stems	-o	-i	-e	-e	-en
-t- and *-l-* stems	-am	-i	-e	-o	-en

The most interesting area of Bengali morphology is the derivation of inflecting stems from verbal bases. Properly speaking, a formal analysis of Bengali verbal stem derivation presupposes the statement of various morphophonological rules. However, for the sake of brevity and clarity, the phenomena will be outlined below more or less informally.

But before the system of verbal stem derivational marking can be discussed, two facts must be presented concerning the shapes of Bengali verbal bases, i.e. the bases to which the stem markers are added.

First, Bengali verbal bases are all either monosyllabic (such as *jan-* 'know') or disyllabic (such as *kamṛa-* 'bite'). The first syllabic in the verbal base may be called the root vowel. There is a productive process for deriving disyllabic bases from monosyllabics by the addition of a stem vowel. This stem vowel is *-a-* (post-vocalically *-oa-*) as in *jana-* 'inform'; although, for many speakers, the stem vowel may be *-o-* if the root vowel (i.e. of the monosyllabic base) is [+high]; e.g. *jiro-*, for some speakers *jira-* 'rest'. Derived disyllabics usually serve as the formal causatives of their monosyllabic counterparts. Compare: *jan-* 'know', *jana-* 'inform'; *oṭh-* 'rise', *oṭha-* 'raise'; *dækh-* 'see', *dækha-* 'show'.

Second, monosyllabic bases with non-high root vowels have two alternate forms, respectively called low and high. Examples are:

	Low alternate base	*High alternate base*
'know'	jan-	jen-
'see'	dækh-	dekh-
'sit'	bɔs-	bos-
'buy'	ken-	kin-
'rise'	oṭh-	uṭh-

When the root vowel is /a/, /e/ is substituted to derive the high alternate base; for bases with front or back non-high root vowels, the high alternate base is formed by assimilating the original root vowel to the next higher vowel in the vowel inventory (see again table 23.2). The latter behaviour suggests an extended application of the Vowel Height Assimilation process discussed in the preceding section. It is, in fact, feasible to state the rules of verb stem derivation so that the low/high alternation is phonologically motivated; i.e. by positing a high vowel (specifically, /i/) in the underlying shapes of the stem-deriving markers. In some verbal forms there is concrete evidence for the /i/ element, as will be observed below. Also, Vowel Height Assimilation must be invoked in any case to account for the fact that, in the derivation of verbal forms which have zero marking of the stem (that is, the present imperative and unmarked (present) indicative), the high alternate base occurs before any inflection containing a high vowel. Thus *dækh-* 'see', *dækho* 'you (ordinary) see', but *dekhi* 'I see', *dekhis* 'you (despective) see', *dekhun* (honorific) 'see!', etc. That there is no high-low alternation in these inflections for disyllabic bases is consistent with the fact that Vowel Height Assimilation only applies when a high syllabic occurs in the immediately succeeding syllable. Thus *oṭha-* 'raise (cause to rise)', *oṭhae* 'he/she raises', *oṭhai* (**uṭhai*) 'I/we raise', etc.

The left-hand column of table 23.4 lists the various Bengali verbal stem types. Two of the verbal forms with Ø stem marking, the present imperative and present indicative, were just discussed. It may be pointed out that, in this stem type, the vowel element /u/ of the third person ordinary inflection *-uk* and of the second/third person honorific inflection *-un*, as well as the /i/ of the second person despective inflection *-is*, all disappear post-vocalically (after Vowel Height Assimilation applies); thus (as above) *dekhis* 'you (despective) see' but (from *hɔ-* 'become') *hok* 'let him/her/it/them become!'; *hon* 'he/she/you/they (honorific) become!'; *hos* 'you (despective) become'.

A verbal form with Ø stem marking not so far discussed is the denominative verbal form or verbal noun. The verbal noun is a non-inflecting form and is therefore not listed in table 23.4. In monosyllabic bases, the marker of this form is suffixed *-a* (*-oa* post-vocalically); for most standard dialect speakers, the marker in disyllabics is *-no*. Thus *oṭh-* 'rise', *oṭha* 'rising', *oṭha-* 'raise', *oṭhano* 'raising'; *jan-* 'know', *jana* 'knowing', *jana-* 'inform', *janano* 'informing'; *ga-* 'sing', *gaoa* 'singing', *gaoa-* 'cause to sing', *gaoano* 'causing to sing'.

Continuing in the leftmost column of table 23.4, the stem-deriving marker *-(c)ch-* signals continuative aspect and is used, independent of any other derivational marker, to derive the present continuous verbal form. The element (*c*) of the marker *-(c)ch-* deletes post-consonantally; compare *khacche* 'is eating' (from *kha-*) with *anche* 'is bringing' (from *an-*). In forming the verbal stem with *-(c)ch-* the high alternate base is selected, unless the base is disyllabic or is a monosyllabic base having the root vowel

/a/. Compare the last examples with *uṭhche* 'is rising' (from *oṭh-*), *oṭhacche* 'is raising' (from *oṭha-*). In a formal treatment of Bengali morphophonemics, the basic or underlying form of the stem marker could be given as *-i(c)ch-*; in this event, one would posit a rule to delete the element /i/ after Vowel Height Assimilation applies, except in a very limited class of verbs including *ga-* 'sing', *sɔ-* 'bear' and *ca-* 'want'. In forming the present continuous forms of these verbs, the element /i/ surfaces, although the element (*c*) of the stem marker tends to be deleted. The resulting shapes are, respectively: *gaiche* 'is singing' (*gacche* is at best non-standard); *soiche* (**socche*) 'is bearing'; *caiche* 'is wanting' (*cacche* does, however, occur as a variant).

The stem-deriving marker *-b-* (see table 23.4) signals irrealis aspect and is used to derive future verbal forms, both indicative and imperative (except for the imperative of the second person ordinary, which will be treated after the next paragraph). In Bengali, the future imperative, as well as the present imperative, may occur in affirmative commands; however, the future imperative, never the present imperative, occurs in negative commands.

In forming the verbal stem with *-b-*, the high alternate base is selected except in three cases: where the base is disyllabic, where the monosyllabic base has the root vowel /a/ and where the monosyllabic base is vowel-final. Thus: *uṭhbo* 'I/we will rise' (from *oṭh-*), but *oṭhabo* 'I/we will raise' (from *oṭha-*); *janbo* 'I/we will know' (from *jan-*), *debo* 'I/we will give' (from *de-*). Compare, however, *dibi* 'you (despective) will give', where Vowel Height Assimilation raises the root vowel. It is possible, again, to posit an underlying /i/ in the irrealis stem marker's underlying shape (i.e. *-ib-*), with deletion of the element /i/ applying except for the small class of verbs noted earlier; thus *gaibo* (**gabo*) 'I/we will sing', *soibo* (**sobo*) 'I/we will bear', *caibo* (**cabo*) 'I/we will want'.

The future imperative of the second person ordinary takes the termination *-io*, which can be analysed as a stem formant *-i-* followed by the second person ordinary inflection *-o* (which is also added to unmarked stems, as table 23.4 shows). When combining with this marker *-i-*, all monosyllabic bases occur in their high alternate shapes; e.g. *hoio* 'become!' (from *hɔ-*). The *-i-* marker is deleted post-consonantally, hence *uṭho* 'rise!' (from *oṭh-*); it also deletes when added to most monosyllabic bases terminating in final /a/, for instance: *peo* 'get!' (**peio*) (from *pa-* 'receive'); *geo* 'sing!' (from *ga-* 'sing'). Bengali disyllabic bases drop their final element /a/ or /o/ before the future imperative stem marker *-i-*. Vowel Height Assimilation applies, hence *uṭhio* 'you must raise!' (from *oṭha-*), *dekhio* 'you must show!' (from *dækha-*), *kamṛio* 'you must bite!' (from *kamṛa-*).

Continuing in the left-hand column of table 23.4, the stem-deriving marker *-t-* signals non-punctual aspect and appears in several forms of the Bengali verb. The Bengali infinitive termination is invariant *-te*, e.g. *jante* 'to know' (from *jan-*) (as in *jante cai* 'I want to know'). The marker *-t-* also occurs in the finite verbal form used to express the past habitual and perfect

conditional, e.g. *jantam* 'I/we used to know' or 'if I/we had known'. The high alternate of monosyllabic bases cooccurs with this marker except in those bases containing a root vowel /a/ followed by a consonant. To illustrate, the infinitive of *oṭh-* 'rise' is *uṭhte*; of *oṭha-* 'raise', *oṭhate*; of *de-* 'give', *dite*; of *hɔ-* 'become', *hote*; of *kha-* 'eat', *khete*; of *an-* 'bring', *ante* (**ente*). Similarly, *uṭhtam* 'I/we used to rise' or 'if I/we had risen'; *oṭhatam* 'I/we used to raise' or 'if I/we had raised', etc. As before, evidence for an /i/ element in the underlying form of the marker *-t-* (i.e. *-it-*) comes from the earlier noted class of verbs 'sing', etc.; for example, *gaite* (**gate*) 'to sing', *gaitam* (**gatam*) 'I/we used to sing' or 'if I/we had sung'; *soite* (**sote*) 'to bear', *soitam* (**sotam*) 'I/we used to bear' or 'if I/we had borne'; *caite* (**cate*) 'to want', *caitam* (**catam*) 'I/we used to want' or 'if I/we had wanted', etc.

The stem-deriving marker *-l-* signals anterior aspect and appears in two verbal forms. The termination of the imperfect conditional is invariant *-le*, e.g. *janle* 'if one knows' (from *jan-*). The marker *-l-* also occurs in the ordinary past tense verbal form, e.g. *janlam* 'I/we knew'. The behaviour of monosyllabic verbal bases in coocurrence with this marker is the same as their behaviour in coöccurrence with the marker *-t-* discussed above. Thus *uṭhle* 'if one rises', *oṭhale* 'if one raises', *dile* 'if one gives', *hole* 'if one becomes', *khele* 'if one eats', *anle* 'if one brings'; *uṭhlam* 'I/we rose', *oṭhalam* 'I/we raised'; and, again, *gaile* (**gale*) 'if one sings', *soile* (**sole*) 'if one bears', *caile* (**cale*) 'if one wants'; *gailam* 'I/we sang', and so on.

To complete the account of the conjugation of the Bengali verb it is only necessary to mention that certain stem-deriving markers can be combined on a single verbal base. For instance, the marker *-l-* combined with the uninflected stem in *-(c)ch-* yields a verbal form called the past continuous. Illustrations are: *uṭhchilam* 'I was/we were rising' (from *oṭh-*), *oṭhacchilam* 'I was/we were raising' (from *oṭha-*), *khacchilam* 'I was/we were eating' (from *kha-*).

It is also possible to combine stem-deriving markers on the Bengali verbal base in the completive aspect. The marker of this aspect is *-(i)e-*, not listed in table 23.4 because it is not used in isolation from other stem-forming markers to form inflecting verbal stems. Independently of any other stem-forming marker it may, however, be added to a verbal base to derive a non-finite verbal form known as the conjunctive participle (or gerund). An example is: *bujhe* 'having understood' from *bujh-* 'understand' (note that the element (*i*) of *-(i)e-* deletes post-consonantally). When attached to the completive aspect marker *-(i)e-*, all monosyllabic bases occur in their high alternate shapes; disyllabic bases drop their final element /a/ or /o/; and in the latter case, Vowel Height Assimilation applies. Thus: *uṭhe* 'having risen' (from *oṭh-*); *jene* 'having known' (from *jan-*); *diye* 'having given' (from *de-*); *uṭhie* 'having raised' (from *oṭha-*), *janie* 'having informed' (from *jana-*). Now the stem-deriving marker *-(c)ch-* may combine with the verbal stem in *-(i)e-*, yielding a verbal form called the present perfect; the combining shape of the

Bengali Verbal Conjugation Types

	pa- 'receive'		*an-* 'bring'		*bɔs-* 'sit'		*bɔsa-* 'seat'	
Verbal noun	paoa	'receiving'	ana	'bringing'	bɔsa	'sitting'	bɔsano	'seating'
Present indicative	pae	'receives'	ane	'brings'	bɔse	'sits'	bɔsae	'seats'
Present imperative	pak	'let him/her/them receive!'	anuk	'let him/her/them bring!'	bosuk	'let him/her/them sit!'	bɔsak	'let him/her/them seat!'
Present continuous	pacche	'is receiving'	anche	'is bringing'	bosche	'is sitting'	bɔsacche	'is seating'
Future indicative/future imperative	pabe	'will receive'/'must receive!'	anbe	'will bring'/'must bring!'	bosbe	'will sit'/'must sit!'	bɔsabe	'will seat'/'must seat!'
Infinitive	pete	'to receive'	ante	'to bring'	boste	'to sit'	bɔsate	'to seat'
Perfect conditional/past habitual	peto	'would receive'	anto	'would bring'	bosto	'would sit'	bɔsato	'would seat'
Imperfect conditional	pele	'if one receives'	anle	'if one brings'	bosle	'if one sits'	bɔsale	'if one seats'
Ordinary past	pelo	'received'	anlo	'brought'	boslo	'sat'	bɔsalo	'seated'
Past continuous	pacchilo	'was receiving'	anchilo	'was bringing'	boschilo	'was sitting'	bɔsacchilo	'was seating'
Conjunctive participle	peye	'having received'	ene	'having brought'	bose	'having sat'	bosie	'having seated'
Present perfect	peyeche	'has received'	eneche	'has brought'	boseche	'has sat'	bosieche	'has seated'
Past perfect	peyechilo	'had received'	enechilo	'had brought'	bosechilo	'had sat'	bosiechilo	'had seated'

former marker in such cases is invariably *-ch-*. This is to say that the element *(c)* of the marker *-(c)ch-* not only deletes post-consonantally (see the earlier discussion of continuous aspect marking), but also following the stem-deriving marker *-(i)e-*. Some examples are: *dekheche* 'has seen' (from monosyllabic *dækh-*), *dekhieche* 'has shown' (from disyllabic *dækha-*), *diyeche* 'has given' (from *de-* 'give'). The verbal stem in *-(i)e-* followed by *-(c)ch-* may further combine with the anterior aspect marker *-l-* to yield a verbal form called the past perfect; e.g. *dekhechilam* 'I/we had seen', *dekhiechilam* 'I/we had shown'.

Examples of conjugation for four Bengali verbal bases are given in the chart of verbal conjugation types. The inflection illustrated in the chart is the third person ordinary.

4 Syntax

The preceding discussion of declensional parameters (case and number for nouns, person and status for verbs) ties in naturally with the topic of agreement in Bengali syntax. A number of modern Indo-Aryan languages (see, for example, the chapter on Hindi-Urdu) demonstrate a degree of ergative patterning in predicate-noun phrase agreement; and Bengali, in its early historical stages, likewise showed some ergative patterning (i.e. sentential verb agreeing with subject of an intransitive sentence but with object, not subject, of a transitive sentence). However, this behaviour is not characteristic today of any of the eastern Magadhan languages.

Thus in Modern Bengali, sentences normally have subjects in the nominative or unmarked case, and the finite predicates of sentences normally agree with their subjects for the parameters of person and status. There are, however, two broad classes of exceptions to this generalisation. The passive constructions exemplify one class. Passive in Modern Bengali is a special variety of sentence nominalisation. When a sentence is nominalised, the predicate takes the verbal noun form (discussed in the preceding section) and the subject is marked with the genitive case. Under passivisation, a sentence is nominalised and then assigned to one of a small set of matrix predicates, the most common being *hɔ-* 'become' and *ja-* 'go'; and when the latter is selected, the subject of the nominalised sentence is obligatorily deleted. Examples are: *tomar jɔthesṭo khaoa hoyeche?* (your enough eating has-become) 'have you eaten enough?' (i.e. has it been sufficiently eaten by you?) and *oke paoa gælo* (to-him getting it-went) 'he was found' (i.e. him was found). In a passive sentence, the matrix verb (*hɔ-* or *ja-*) lacks agreement with any noun phrase. In particular, it cannot agree with the original subject of the active sentence — this noun phrase has become marked with the genitive case under nominalisation, or deleted altogether. This is to say that the Modern Bengali passive construction lacks a formal subject; it is of a type referred to in some grammatical literature as

the 'impersonal passive'. These constructions form one class of exceptions to the characteristic pattern of Bengali subject-verb agreement.

The other class of exceptions comprises certain expressions having subjects which occur in a marked or oblique case. In Bengali there are a few complex constructions of this type. Bengali also has several dozen predicates which regularly occur in non-complex constructions with marked subjects. These constructions can be called indirect subject constructions, and indirect subjects in Modern Bengali are invariably marked with the genitive case. (At an earlier historical stage of the language, any of the oblique cases could be used for the marking of the subject in this sort of construction.) In the Modern Bengali indirect subject construction, the finite predicate normally demonstrates no agreement. An example is: *maer tomake pɔchondo hɔy* (of-mother to-you likes) 'Mother likes you'. Bengali indirect subject predicates typically express sensory, mental, emotional, corporal and other characteristically human experiences. These predicates constitute a significant class of exceptions to the generalised pattern of subject-finite predicate agreement in Modern Bengali.

The remainder of this overview of Bengali syntax will be devoted to the topic of word order, or the relative ordering of major constituents in sentences. In some literature on word order types, Bengali has been characterised as a rigidly verb-final language, wherein nominal modifiers precede their heads; verbal modifiers follow verbal bases; the verbal complex is placed sentence-finally; and the subject noun phrase occupies the initial position in a sentence. In these respects Bengali is said to contrast with earlier Indo-Aryan, in which the relative ordering of sentential constituents was freer, notwithstanding a statistical tendency for verbs to stand at the ends of their clauses.

It is true that the ordering of sentential elements is more rigid in Modern Bengali than in Classical Sanskrit. However, the view that Bengali represents a 'rigid' verb-final language does not adequately describe its differences from earlier Indo-Aryan word order patterning.

Word order within the Modern Bengali noun phrase is, to be sure, strict. An adjective or genitive expression is always placed before the noun it modifies. By contrast, in earlier Indo-Aryan, adjectives showed inflectional concord with their modified nouns and consequently were freer in their positioning; more or less the same applied to the positioning of genitive expressions with respect to nominal heads. Not only is the ordering of elements within the noun phrase more rigid in Modern Bengali, but the mutual ordering of noun phrases within the sentence is strict as well, much more so than in earlier Indo-Aryan. The subject noun phrase generally comes first in a Modern Bengali sentence, followed by an indirect object if one occurs; next comes the direct object if one occurs; after which an oblique object noun phrase may be positioned. This strictness of linear ordering can be ascribed to the relative impoverishment of the Modern Bengali case

system in comparison with earlier Indo-Aryan. Bengali case markers are, nonetheless, supplemented by a number of postpositions, each of which may govern nouns declined in one of two cases, the objective or genitive.

We will now consider word order within the verb phrase. At the Old Indo-Aryan stage exemplified by Classical Sanskrit, markers representing certain verbal qualifiers (causal, desiderative, potential and conditional) could be affixed to verbal bases, as stem-forming markers and/or as inflectional endings. Another verbal qualifier, the marker of sentential negation, tended to be placed just before the sentential verb. The sentential interrogative particle, on the other hand, was often placed at a distance from the verbal complex.

In Modern Bengali, the only verbal qualifier which is regularly affixed to verbal bases is the causal. (See the discussion of derived disyllabic verbal bases in section 3 above.) The following pair of Bengali sentences illustrates the formal relationship between non-causative and causative constructions: *cheleṭi ciṭhiṭa poṛlo* (the-boy the-letter read) 'the boy read the letter'; *ma cheleṭi-ke diye ciṭhiṭa pɔṛalen* (mother to-the-boy by the-letter caused-to-read) 'the mother had the boy read the letter'. It will be noted that in the second example the non-causal agent is marked with the postposition *diye* 'by' placed after its governed noun, which appears in the objective case. Usually, when the verbal base from which the causative is formed is transitive, the non-causal agent is marked in just this way. The objective case alone is used to mark the non-causal agent when the causative is derived either from an intransitive base, or from any of several semantically 'affective' verbs — transitive verbs expressing actions whose principal effect accrues to their agents and not their undergoers. Examples are: 'eat', 'smell', 'hear', 'see', 'read' (in the sense of 'study'), 'understand' and several others.

It was mentioned above that the modalities of desiderative and potential action could be marked on the verbal form itself in Old Indo-Aryan. In Modern Bengali, these modalities are usually expressed periphrastically; i.e. by suffixing the infinitive marker to the verbal stem, which is then followed by a modal verb. To illustrate: *uṭhte cae* 'wants to rise', *uṭhte pare* 'can rise'.

Conditional expressions occur in two forms in Modern Bengali. The conditional clause may be finite, in which case there appears the particle *jodi*, which is a direct borrowing from a functionally similar Sanskrit particle *yadi*. To illustrate: *jodi tumi kajṭa sarbe (tɔbe) eso* (if you the-work will-finish (then) come) 'if/when you finish the work, (then) come over!'. An alternate way of framing a conditional is by means of the non-finite conditional verbal form (imperfect conditional), which was mentioned in section 3. In this case no conditional particle is used; e.g. *tumi kajṭa sarle (tɔbe) eso* (you the-work if-finish (then) come) 'if/when you finish the work, come over!'.

The particle of sentential negation in Bengali is *na*. In independent clauses it generally follows the sentential verb; in subjoined clauses (both finite and non-finite), it precedes. Thus: *boslam na* (I-sat not) 'I did not sit'; *jodi tumi na bɔso* (if you not sit) 'if you don't sit'; *tumi na bosle* (you not if-sit) 'if you don't sit'. Bengali has, it should be mentioned, two negative verbs. Each of them is a counterpart to one of the verbs 'to be'; and in this connection it needs to be stated that Bengali has three verbs 'to be'. These are respectively the predicative *hɔ-* 'become'; the existential verb 'exist', having independent/subjoined clause allomorphs *ach-/thak-*; and the equational verb or copula, which is normally Ø but in emphatic contexts is represented by *hɔ-* placed between two arguments (compare, for example, non-emphatic *ini jodu* (this-person Ø Jodu) 'this is Jodu' versus emphatic *ini hocchen jodu* (this-person is Jodu) 'this (one) is Jodu'). While the predicative verb 'to be' has no special negative counterpart (it is negated like any other Bengali verb), the other two verbs 'to be' each have a negative counterpart. Moreover, for each of these negative verbs, there are separate allomorphs which occur in independent and subjoined clauses. The respective independent/subjoined shapes of the negative verbs are existential *nei/na thak-* (note that the verb *nei* is invariant) and equational *nɔ-/na hɔ-*. It bears mentioning, incidentally, that negative verbs are neither characteristic of modern nor of earlier Indo-Aryan. They are, if anything, reminiscent of negative copulas and other negative verbs in languages of the Dravidian (South Indian) family, such as Modern Tamil.

The Modern Bengali sentential interrogative particle *ki* is inherited from an earlier Indo-Aryan particle of similar function. The sentential interrogative *ki* may appear in almost any position in a Bengali sentence other than absolute initial; however, sentences vary in their presuppositional nuances according to the placement of this particle, which seems to give the most neutral reading when placed in the second position (i.e. after the first sentential constituent). To illustrate, compare: *tumi ki ekhane chatro?* (you interrogative here student) 'are you a student here?'; *tumi ekhane ki chatro?* (you here interrogative student) 'is it here that you are a student?'; *tumi ekhane chatro (na) ki?* (you here student (negative) interrogative) 'oh, is it that you are a student here?'.

To complete this treatment of word order, we may discuss the relative ordering of marked and unmarked clauses in Bengali complex sentences. By 'marked clause' is meant either a non-finite subordinate clause or a clause whose function within the sentential frame is signalled by some distinctive marker; an instance of such a marker being *jodi*, the particle of the finite conditional clause. As a rule, in a Bengali sentence containing two or more clauses, marked clauses tend to precede unmarked. This is, for instance, true of conjunctive participle constructions; e.g. *bari giye kapoṛ cheṛe ami can korlam* (home having-gone clothes having-removed I bath did) 'going home and removing my clothes, I had a bath'. Relative clauses in Bengali likewise generally precede main clauses, since they are marked (that is, with relative pronouns); Bengali, then, exhibits the correlative sentential type

which is well attested throughout the history of Indo-Aryan. An illustration of this construction is: *je boiṭa enecho ami seṭa kichu din rakhbo* (which book you-brought I it some days will-keep) 'I shall keep the book you have brought for a few days'. Finite complement sentences marked with the complementiser *bole* (derived from the conjunctive participle of the verb *bɔl-* 'say') likewise precede unmarked clauses; e.g. *apni jacchen bole ami jani* (you are-going complementiser I know) 'I know that you are going'.

An exception to the usual order of marked before unmarked clauses is exemplified by an alternative finite complement construction. Instead of clause-final marking (with *bole*), the complement clause type in question has an initial marker, a particle *je* (derived historically from a complementiser particle of earlier Indo-Aryan). A complement clause marked initially with *je* is ordered invariably after, not before, the unmarked clause; e.g. *ami jani je apni jacchen* (I know complementiser you are-going) 'I know that you are going'.

5 Concluding Points

In this final section the intention is to relate the foregoing discussion to the question of Bengali's historical development and present standing, both within the Indo-Aryan family and within the general South Asian language area. To accomplish this, it is useful to consider the fact of lectal differentiation in the present community of Bengali speakers. Both vertical and horizontal varieties are observed.

Vertical differentiation, or diglossia, is a feature of the current standard language. This is to say that the language has two styles used more or less for complementary purposes. Of the two styles, the literary or 'pundit language' (*sadhu bhasa*) shows greater conservatism in word morphology (i.e. in regard to verbal morphophonemics and the shapes of case endings) as well as in lexis (it is characterised by a high frequency of words whose forms are directly borrowed from Sanskrit). The less conservative style identified with the spoken or 'current language' (*colti bhasa*) is the everyday medium of informal discourse. Lately it is also gaining currency in more formal discourse situations and, in written expression, has been encroaching on the literary style for some decades.

The institutionalisation of the *sadhu-colti* distinction occurred in Bengali in the nineteenth century, and (as suggested in the last paragraph) shows signs of weakening today. Given (1) that the majority of Bengali speakers today are not Hindu and cannot be expected to maintain an emotional affinity to Sanskritic norms, plus (2) the Bangladesh government's recent moves to enhance the Islamic character of eastern Bengali society and culture and (3) the fact that the colloquial style is overtaking the literary even in western Bengal (both in speech and writing), it remains to be seen

over the coming years whether a formal differentiation of everyday versus 'pundit' style language will be maintained.

It should be added that, although throughout the Bengali-speaking area a single, more or less uniform variety of the language is regarded as the standard dialect, the bulk of speakers have at best a passing acquaintance with it. That is, horizontal differentiation of Bengali lects is very extensive (if poorly researched), both in terms of the number of regional dialects that occur and in terms of their mutual divergence. (The extreme eastern dialect of Chittagong, for instance, is unintelligible even to many speakers of other eastern Bengali dialects.) The degree of horizontal differentiation that occurs in the present Bengali-speaking region is related to the ambiguity of Bengali's linguistic affiliation, i.e. areal as contrasted with genetic. It is to be noted that the Bengali-speaking region of the Indian subcontinent to this day borders on or subsumes the domains of a number of non-Indo-Aryan languages. Among them are Malto (a Dravidian language of eastern Bihar); Ahom (a Tai language of neighbouring Assam); Garo (a Tibeto-Burman language spoken in the northern districts of Bengal itself); as well as several languages affiliated with Munda (a subfamily of Austro-Asiatic), such as Santali and Mundari (both of these languages are spoken within as well as outside the Bengali-speaking area).

It has been pointed out earlier that modern standard Bengali has several features suggestive of extra-Aryan influence. These features are: the frequency of retroflex consonants; initial-syllable word stress; absence of grammatical gender; negative verbs. Though not specifically pointed out as such previously, Bengali has several other formal features, discussed above, which represent divergences from the norms of Indo-Aryan and suggest convergence with the areal norms of greater South Asia. These features are: post-verbal negative particle placement; clause-final complement sentence marking; relative rigidity of word order patterning in general, and sentence-final verb positioning in particular; proliferation of the indirect subject construction (which was only occasionally manifested in early Indo-Aryan).

In addition to the above, it may be mentioned that Bengali has two lexical features of a type foreign to Indo-Aryan. These features are, however, not atypical of languages of the general South Asian language area (and are even more typical of South-East Asian languages). One of these is a class of reduplicative expressives, words such as: *kickic* (suggesting grittiness), *miṭmiṭ* (suggesting flickering), *ṭɔlmɔl* (suggesting an overflowing or fluid state). There are dozens of such lexemes in current standard Bengali. The other un-Aryan lexical class consists of around a dozen classifier words, principally numeral classifiers. Examples are: *du jon chatro* (two human-classifier student) 'two students'; *tin khana boi* (three flat-thing-classifier book) 'three books'.

It is probable that the features discussed above were absorbed from other languages into Bengali after the thirteenth century, as the language came to

be increasingly used east of the traditional sociocultural centre of Bengal. That centre, located along the former main course of the Ganges (the present-day Bhagirathi-Hooghley River) in western Bengal, still sets the standard for spoken and written expression in the language. Thus standard Bengali is defined even today as the dialect spoken in Calcutta and its environs. It is a reasonable hypothesis nevertheless, as suggested above in section 1, that descendants of non-Bengali tribals of a few centuries past now comprise the bulk of Bengali speakers. In other words, the vast majority of the Bengali linguistic community today represents present or former inhabitants of the previously uncultivated and culturally unassimilated tracts of eastern Bengal. Over the past several centuries, these newcomers to the Bengali-speaking community are the ones responsible for the language's having acquired a definite affiliation within the South Asian linguistic area, above and beyond the predetermined and less interesting fact of its genetic affiliation in Indo-Aryan.

Bibliography

Chatterji (1926) is the classic, and indispensable, treatment of historical phonology and morphology in Bengali and the other Indo-Aryan languages. A good bibliographical source is Čižikova and Ferguson (1969). For the relation between literary and colloquial Bengali, see Dimock (1960).

The absence of a comprehensive reference grammar of Bengali in English is noticeable. Ray et al. (1966) is one of the better concise reference grammars. Chatterji (1939) is a comprehensive grammar in Bengali, while Chatterji (1972) is a concise but thorough treatment of Bengali grammar following the traditional scheme of Indian grammars. Two pedagogical works are also useful: Dimock et al. (1965), a first-year textbook containing very lucid descriptions of the basic structural categories of the language, and Bender and Riccardi (1978), an advanced Bengali textbook containing much useful information on Bengali literature and on the modern literary language. For individual topics, the following can be recommended: on phonetics-phonology, Chatterji (1921) and Ferguson and Chowdhury (1960); on the morphology of the verb, Dimock (1957), Ferguson (1945) and Sarkar (1976); on syntax, Klaiman (1981), which discusses the syntax and semantics of the indirect subject construction, passives and the conjunctive participle construction in modern and earlier stages of Bengali.

References

Bender, E. and T. Riccardi, Jr. 1978. *An Advanced Course in Bengali* (South Asia Regional Studies, University of Pennsylvania, Philadephia)

Chatterji, S.K. 1921. 'Bengali Phonetics', in *Bulletin of the School of Oriental and African Studies*, vol. 2, pp. 1–25

——— 1926. *The Origin and Development of the Bengali Language*, 3 vols. (Allen and Unwin, London)

——— 1939. *Bhāṣāprakāśa bāṅgālā byākaraṇa* (Calcutta University, Calcutta)

——— 1972. *Sarala bhāṣāprakāśa bāṅgālā byākaraṇa*, revised ed. (Bāk-sāhitya, Calcutta)

Čižikova, K.L. and C.A. Ferguson. 1969. 'Bibliographical Review of Bengali Studies', in T. Sebeok (ed.), *Current Trends in Linguistics*, vol. 5: *Linguistics in South Asia* (Mouton, The Hague), pp. 85–98

Dimock, E.C., Jr. 1957. 'Notes on Stem-vowel Alternation in the Bengali Verb', *Indian Linguistics*, vol. 17, pp. 173–7

——— 1960. 'Literary and Colloquial Bengali in Modern Bengali Prose', *International Journal of American Linguistics*, vol. 26, no. 3, pp. 43–63

——— et al. 1965. *Introduction to Bengali*, part 1 (East-West Center, Honolulu; reprinted South Asia Books, Columbia, Mo., 1976)

Ferguson, C.A. 1945. 'A Chart of the Bengali Verb', *Journal of the American Oriental Society*, vol. 65, pp. 54–5

——— and M. Chowdhury. 1960. 'The Phonemes of Bengali', *Language*, vol. 36, pp. 22–59

Klaiman, M.H. 1981. *Volitionality and Subject in Bengali: A Study of Semantic Parameters in Grammatical Processes* (Indiana University Linguistics Club, Bloomington)

Ray, P.S., et al. 1966. *Bengali Language Handbook* (Center for Applied Linguistics, Washington, DC)

Sarkar, P. 1976. 'The Bengali Verb', *International Journal of Dravidian Linguistics*, vol. 5, pp. 274–97

24 Iranian Languages

J.R. Payne

The approximate present distribution of the Iranian languages is illustrated in the attached sketch-map. The languages currently spoken, according to their genetic relations within Iranian (see below) are:

South-West Iranian: Persian (Iran, Persian Gulf); Dari (Afghanistan); Tajiki (USSR); Luri and Bakhtiari (nomadic, Iran); Kumzari (Persian Gulf); non-Persian dialects of Fars province, centred on Shiraz, Kazerun, Sivand and Lar (Iran); Tati (USSR).

North-West Iranian: Kurdish (Turkey, Iran, Iraq, Syria, USSR); Talishi (USSR, Iran); Balochi (Pakistan, Iran, Afghanistan, USSR, Persian Gulf); Gilaki (Iran); Mazandarani (Iran); Zaza (Turkey); Gurani (Iran, Iraq); Bashkardi (Iran); Parachi (Afghanistan); Ormuri (Afghanistan, Pakistan); Semnani and related dialects (Iran); 'Tat' dialects, centred on Tabriz, Zanjan, Qazvin and Saveh (Iran); Vafsi and Ashtiyani (Iran); dialects of central Iran, centred on Kashan, Esfahan, Yazd, Kerman and the Dashte-Kavir (Iran).

South-East Iranian: Pashto (Afghanistan, Pakistan); Yazgulami (USSR); Shughni (USSR, Afghanistan); Roshani (USSR, Afghanistan); Bartangi (USSR); Oroshori (USSR); Sarikoli (China); Ishkashmi (Afghanistan, USSR); Sanglechi (Afghanistan); Zebaki (Afghanistan); Wakhi (Afghanistan, USSR, Pakistan, China); Munji (Afghanistan); Yidgha (Pakistan).

North-East Iranian: Ossete (USSR), Yaghnobi (USSR).

It will be noted that the names of the genetic groups do not always accurately reflect the geographic location of the modern languages. In particular, Ossete, which belongs to the North-East group, is spoken in the Caucasus,

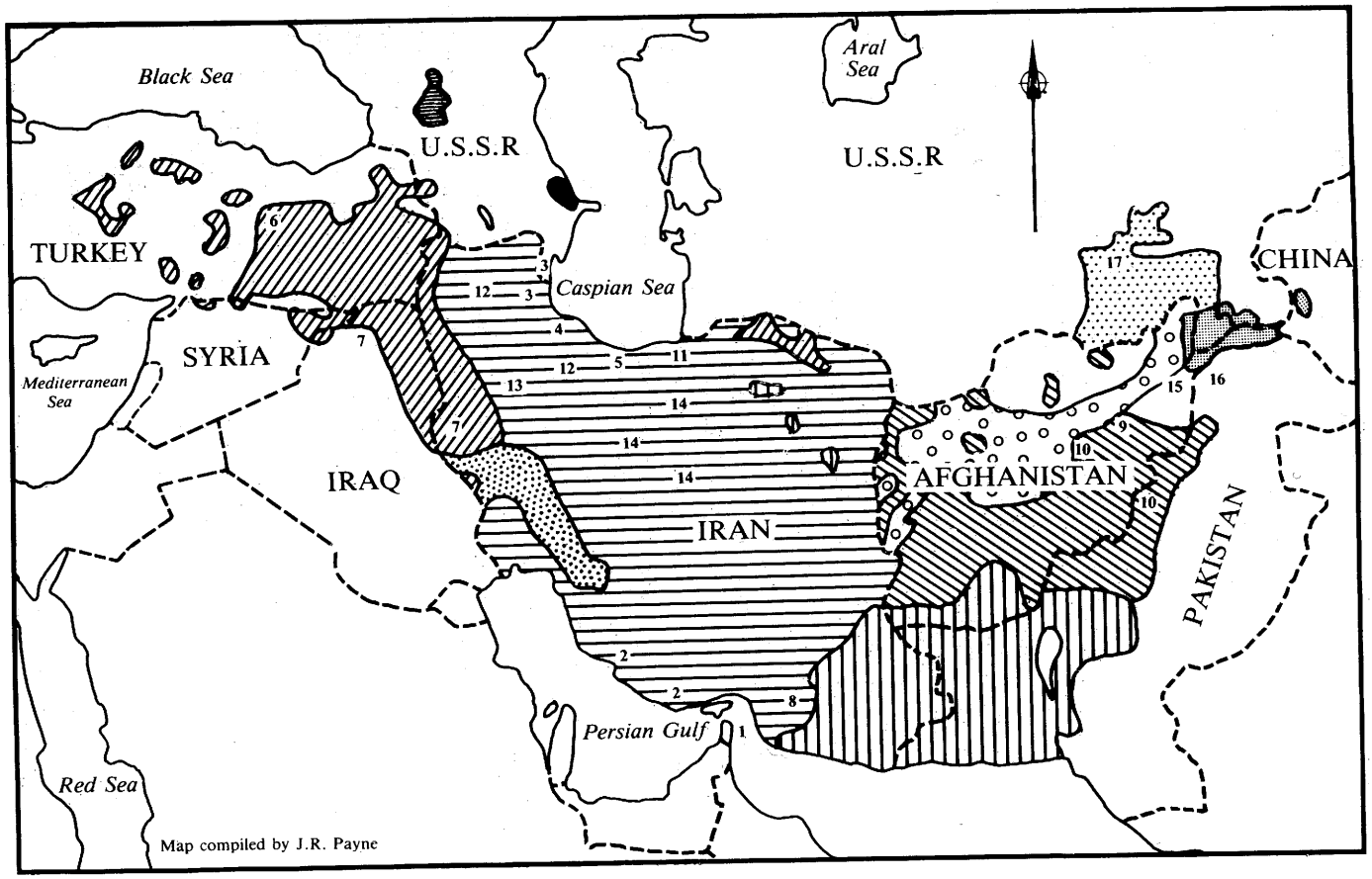
which represents the north-west of the present Iranian language area, and Balochi, which belongs to the North-West group, is located in the extreme south-east on either side of the Iran-Pakistan border. In fact, the geographic nomenclature is more closely tied to the distribution of extinct Iranian languages from the Old Iranian (up to the fourth/third centuries BC) and Middle Iranian (from the fourth/third centuries BC to the eighth/ninth centuries AD) periods.

The oldest attested forms of Iranian are Old Persian, known from the cuneiform inscriptions of the Achaemenid emperors, in particular Darius the Great (521–486 BC) and Xerxes (486–465 BC), and Avestan, the language of the Avesta, a collection of sacred Zoroastrian texts. The oldest parts of the Avesta, the Gathas or songs attributed to the prophet Zoroaster himself, reflect a slightly more archaic stage of development than the Old Persian inscriptions, and must therefore be dated to the sixth century BC or earlier, although the first manuscripts are from the thirteenth and fourteenth centuries AD. Genetically, Old Persian can be clearly associated with the South-West Iranian group, the Achaemenid empire being centred on the province of Persis in the south-west of modern Iran, and must be considered a direct precursor of forms of Middle and Modern Persian. The position of Avestan is, however, complex and disputed, as might be expected of an orally transmitted religious text. The focus of Zoroastrian conversion is held to be Bactria, south of the Oxus river in the east, and the Gathas do indeed show some east Iranian characteristics, notably a tendency to voice clusters which appear as *-ft-* and *-xt-* in West Iranian (see below). A possible explanation for the occurrence of some West Iranian forms is the subsequent spread of Zoroastrianism towards Media in the north-west. It is clear, nevertheless, that Avestan shows none of the features characteristic of South-West Iranian.

From archaeological and textual evidence, it can be deduced that Iranian languages at the time of the Achaemenid empire had a wider geographical distribution than at present, extending from the steppes of southern Russia in the west to areas of Chinese Turkestan (Sinkiang) in the east. Old Persian and Avestan are the main linguistic sources from this period; however, proper names and toponyms provide some information about Median, the language of the province of Media centred on Ecbatana (modern Hamadan in north-west Iran), and about the language of the Scythian and Sarmatian tribes of the south Russian steppes. The Median dialect, which belongs genetically to the North-West group, was originally the language of the Median empire (eighth to sixth centuries BC), and some of its influence can be seen in the Old Persian inscriptions. Knowledge of the Scythian and Sarmatian dialects is based on the analysis of proper names and toponyms in inscriptions from the Greek colonies of the period and by comparison with forms of Ossete, the only modern descendant.

By comparison, the Middle Iranian period provides a wealth of materials.

Map 24.1: Approximate Distribution of Iranian Languages



Persian

Dari

Tajiki

Kurdish

Pashto

Luri and Bakhtiari

Balochi

Ossete

Tati

Pamir Languages*

* Shughni, Roshani, Bartangi, Oroshori, Sarikoli, Yazgulami, Wakhi, Ishkashmi, Sanglechi, Zebaki

1 Kumzari
2 Dialects of Fars Province
3 Talishi
4 Gilaki
5 Mazandarani
6 Zaza
7 Gurani
8 Bashkardi
9 Parachi
10 Ormuri
11 Semnani
12 'Tat' dialects
13 Vafsi and Ashtiyani
14 Dialects of Central Iran
15 Munji
16 Yidgha
17 Yaghnobi

To the South-West group belongs Middle Persian, the direct descendant of Old Persian and the precursor of Modern Persian. Although the earliest documents, inscriptions on coins, date from the second century BC, the main corpus illustrates the language of the Sassanid empire (third to seventh centuries AD), centred on the province of Fars (ancient Persis), but by the time of the Arab conquest (seventh/eighth centuries AD) extending over a wide area of present-day Iran, Afghanistan and Central Asia. It includes both secular and Zoroastrian documents written in the Pahlavi script, which is based on the Aramaic and does not show short vowels. The term *Pahlavi* itself is the adjective from the noun *Pahlav* < *Parθava* 'Parthia'. Middle Persian is also represented by a large corpus of Manichean texts found in Turfan, Chinese Turkestan (Sinkiang), and dating mainly to the eighth and ninth centuries AD, although the earliest documents go as far back as the time of Mani (AD 216–74), the founder of the religion. These latter are written mostly in the Manichean script, another derivative of Aramaic, but are also found in Sogdian and Runic Turkic forms.

To the North-West group, apart from Median, belongs Parthian, the source of the Middle Persian script. Parthian itself is more sparsely documented than Middle Persian, but was the language of the province of Parthia which flourished at the time of the Arsacid dynasty (third century BC) to the south-east of the Caspian Sea. It is known through Parthian versions of Sassanid inscriptions and Manichean texts, as well as through minor documents from the first century BC and ostraca from ancient Nisa, located near Ashkhabad in modern Soviet Turkmenia.

For the North-East group there are two representatives. Sogdian, the lingua franca of an extensive area centred on Samarkand and the silk route to China, is known in a number of forms and scripts. In the Sogdian script proper are letters from the fourth century AD, an archive of secular documents dating to the eighth century AD from Mt. Mugh in the Zeravshan area of Tajikistan, as well as a number of Buddhist texts of the same period. There is also an extensive corpus of Manichean and Christian texts, some of the latter written in the Syriac script. The modern descendant of Sogdian is Yaghnobi, spoken until very recently by a small group in one of the high valleys of the Zeravshan, but now dissipated to more lowland areas. Also important as a representative of North-East Iranian in the Middle Iranian period is Khwarezmian, located in a region centred on modern Khiva, and attested in documents and inscriptions in a type of Aramaic script dating mainly to the third to eighth centuries AD. Later fragments of Khwarezmian have survived in Islamic texts of the eleventh to fourteenth centuries AD.

Finally, to the South-East group belong Saka, the language of eastern Scythian tribes from Khotan (Chinese Sinkiang), and Bactrian, the language of the Kushan kingdom of Bactria. The former is known through an extensive corpus of Buddhist texts in the Brahmi script, and dating primarily to the fifth to tenth centuries AD, while the latter is represented mainly by an inscription of twenty-five lines in a variant of the Greek script, found at the temple of Surkh Kotal in northern Afghanistan.

Within the Indo-European family, the Iranian languages are satem languages, e.g. Proto-Indo-European *$\hat{k}$ṃtom 'hundred', Avestan *satəm*, and show a very close relationship to the Indo-Aryan (and Dardic) branches. Three common phonological developments separate Iranian and Indo-Aryan from the rest of Indo-European: (1) the collapse of Proto-Indo-European *a, *e, *o, *$ḷ$, *$ṃ$ into a, and correspondingly of *$ā$, *$ē$, *$ō$, *$ḹ$, *$ṃ̄$ into $ā$, e.g. Proto-Indo-European *$dek̂ṃ$ 'ten' > Avestan *dasa*, Sanskrit *dáśa*, but Old Church Slavonic *desętь*, Latin *decem*; (2) the development of Proto-Indo-European *$ə$ into i, e.g. Proto-Indo-European *$pətē(r)$ 'father' > Old Persian *pitā*, Sanskrit *pitā́*, but Latin *pater*; (3) the development of Proto-Indo-European *s into $š$ or $ṣ$ after *i, *u, *r, *k, e.g. Proto-Indo-European *$ueks$- 'grow' > Old Persian and Avestan *vaxš*-, Sanskrit *vakṣ*-, but German *wachs*-, English *wax*; Proto-Indo-European *sed- 'sit' > Old Persian *ni-šad*-, Sanskrit *ni-ṣīd*- (with additional prefix), but Latin *sed*-, English *sit*. In addition, Iranian and Indo-Aryan inherit from Proto-Indo-European strikingly similar verbal conjugations and nominal declensions. Compare for example the following forms of the first person singular pronoun 'I': (a) nominative: Old Persian *adam*, Avestan *azəm*, Sanskrit *ahám*; (b) accusative: Old Persian *mām*, Avestan *mām*, Sanskrit *mā́m*; (c) genitive: Old Persian *manā*, Avestan *mana*, Sanskrit *máma*; (d) enclitic accusative: Old Persian *-mā*, Avestan *-mā*, Sanskrit *-mā*; (e) enclitic genitive: Old Persian *-maiy*, Avestan *-mōi*, Sanskrit *-mē*; (f) enclitic ablative: Old Persian *-ma*, Avestan *-maṯ*, Sanskrit *-mát*.

In total, according to a recent count, the number of isoglosses linking Iranian with Indo-Aryan is 57, compared with 27 between Indo-Aryan and Greek, 24 between Indo-Aryan and Slavonic and 22 between Indo-Aryan and Baltic. These linguistic facts, in conjunction with shared cultural features such as the name *arya*- 'Aryan', suggest that the Iranian and Indo-Aryan tribes represent a single ethnic and linguistic group within the Indo-European family. Opinions differ, however, as to the dates and routes of migration which led both Iranians and Indo-Aryans away from the Indo-European homeland into the Iranian plateau, Central Asia and India. Since the Rigveda, composed no earlier than the middle of the second millennium BC, already places the Indo-Aryans in India, this date sets a *terminus ante quem* for the loss of Indo-Iranian unity. According to the traditional view, the Aryans must have been in close contact for some time after the break-up of Indo-European, migrating together during the third millennium BC towards Central Asia and the Hindukush. Central Asia then became the focus for the later expansion of Indo-Aryans into India (middle of second millennium BC) and eventually of Iranians into Iran and further west (beginning of first millennium BC). An alternative view, based primarily on archaeological evidence and inscriptions from Mesopotamia, suggests that

the Indo-Aryans split from the Iranians by migrating through the Caucasus at the beginning of the second millennium BC, at a time when both groups were still in contact with other Indo-European groups in southern Russia. Iranian tribes would have maintained this contact, in particular with Greek and Armenian, until they too (at least the western Iranian precursors of the Medes and Persians) entered Iran from the north through the Caucasus at the turn of the first millennium BC.

The main linguistic features characterising the split of Iranian and Indo-Aryan are: (1) Indo-Iranian voiced aspirates **bh*, **dh*, **gh* (< Proto-Indo-European **bh*, **dh*, **gh*) are preserved in Indo-Aryan but converted to *b*, *d*, *g* in Iranian, e.g. Sanskrit *bhrā́tar*, Old Persian and Avestan *brātar*; (2) Indo-Iranian voiceless aspirates **ph*, **th*, **kh* (< Proto-Indo-European **ph*, **th*, **kh* primarily) are preserved in Indo-Aryan but converted to voiceless fricatives *f*, *θ*, *x* in Iranian, or unaspirated stops *p*, *t*, *k* after *s*, e.g. Sanskrit *path-* 'path', Old Persian and Avestan *paθ-*, Sanskrit *sthā-* 'stand', Old Persian and Avestan *stā-*; (3) Indo-Iranian voiceless **p*, **t*, **k* (< Proto-Indo-European **p*, **t*, **k*) become *f*, *θ*, *x* in Iranian when initial in a consonant cluster, e.g. Sanskrit *putrá-* 'son', Avestan *puθra-*, Old Persian *puça-* (with subsequent *θr* > *ç*), but Wakhi, one of the most archaic languages phonologically, preserves the cluster *-tr-*, e.g. *pətr* 'son'; (4) Indo-Iranian palatals **ć*, **ǰ*, **ǰh* (< Proto-Indo-European **k̂*, **ĝ*, **ĝh*) are realised as *s*, *z*, *z* or *θ*, *d*, *d* in Iranian, but *ś*, *j*, *h* in Indo-Aryan, e.g. Sanskrit *hasta-* 'hand', Avestan *zasta-*, Old Persian *dasta-*; (5) Indo-Iranian **s* is preserved in Indo-Aryan, but converted in Iranian, except before **p*, **t*, **k*, into *h*, e.g. Sanskrit *ásmi* '(I) am', Avestan *ahmi*, Old Persian *amiy* (where the *h* is not written). This isogloss *s* > *h*, shared by Greek and Armenian, is used in support of the hypothesis that Iranian tribes entered Iran via the Caucasus rather than from the east.

By the time of the Achaemenids in the middle of the first millenium BC, it is clear that the dialectal divisions are already established which give rise to the modern genetic groupings within Iranian. The basic division between East and West Iranian is characterised by the following correspondences: (1) West Iranian preserves *b*, *d*, *g*, but these are mainly converted in East Iranian into the corresponding voiced fricatives *β* (*v*, *w*), *δ*, *γ*, e.g. Old Persian *brātar* 'brother', Modern Persian *berādar*, Balochi *brās*, but Sogdian *βr't*, Yaghnobi *virōt*; Avestan *dasa* 'ten', Modern Persian *dah*, Bakhtiari *deh*, Zaza *däs*, but Sogdian *δs'*, Shughni *δīs*; Old Persian *gauša* 'ear', Modern Persian *gōš*, Gurani *goš*, Kurdish *goh*, but Sogdian *γwš*, Ossete *γos*, Bartangi *γu*; (2) West Iranian preserves *č*, but this is mainly converted into *c* in East Iranian, e.g. Middle Persian *čahār* 'four', Balochi *čār*, but Khwarezmian *cf'r*/*cβ'r*, Shughni *cavōr*; (3) the consonantal clusters *-ft-* and *-xt-* are preserved in West Iranian, but converted into the voiced counterparts *-vd-* and *-γd-* in East Iranian, equally originally voiced clusters of this type tend to be preserved in East Iranian but devoiced in West

Iranian, e.g. **hafta* 'seven' > Middle Persian *haft*, Kurdish *häft*, but Khwarezmian *'βd*, Ossete *avd*, Yazgulami *uvd*; **duγdar* 'daughter' > Modern Persian *doxtar*, Gilaki *duxtər*, but Avestan *dugədā*, Khwarezmian *δγd*, Wakhi *δəǰd*.

Further phonological characteristics separate the South-West and North-West groups. The South-West Iranian languages, in particular, represent a close-knit group sharing a number of features which distinguish them not only from North-West Iranian but also from East Iranian. The earliest of these, characteristic of the Old Iranian period, is the correspondence North-West, East *s*, *z* = South-West *θ*, *d*, both deriving from the original palatal series (see above), e.g. Avestan *masišta* 'longest', Parthian *msyšt*, but Old Persian *maθišta* 'biggest', Middle Persian *mahist* (with subsequent *θ* > *h*); Avestan *zān-* 'know', Parthian *z'n-*, Gurani *zān-*, Kurdish *zan-*, but Old Persian *dān-*, Modern Persian *dān-*, Tati *dan-*. Later changes *ǰ* > North-West *ǰ-*/*ž-*, South-West *z-*, and *dv-* > North-West *b-*, South-West *d-*, also clearly differentiate the groups, e.g. Parthian *ǰn* 'woman', Zaza *ǰan*, but Middle Persian *zan*, Modern Persian *zan*; Parthian *br* 'door', Zaza *bär*, but Middle Persian *dar*, Modern Persian *dar*. Further subclassification within the North-West group is complicated by the fragmented nature of much of the material and the influence of Persian on many of the dialects, but Gurani and Balochi both preserve archaic characteristics.

Within the East Iranian group, subdivision into South-East and North-East Iranian is based on both phonological and morphological features. The morphological feature characterising the North-East group is the development of a plural marker in *-t* from a suffix originally deriving abstract nouns. Examples of this marker are Sogdian *'wt'k* 'place', plural *'wt'kt*, Yaghnobi *pōda* 'foot', plural *pōdō-t*, and Ossete *sər* 'head', plural *sər-tə*. The South-East group, on the other hand, shows a variety of voiced continuants in place of invervocalic *-š-*, e.g. Yaghnobi *γuš* 'ear', but Shughni *γůǰ*, Munji *γūy*, as well as a tendency to develop retroflex consonants (though these are lacking in the Shughni-Roshani subgroup of Pamir languages). Within the South-East group, Shughni, Roshani, Bartangi, Oroshori and Sarikoli (and more distantly Yazgulami) form a genetic subgroup, as do Ishkashmi, Zebaki and Sanglechi, and Munji and Yidgha. Munji and Yidgha share with Pashto the development of *d* > *l* (see the chapter on Pashto).

All Iranian languages of the Middle and Modern periods exhibit some common characteristics. The unmarked word order is typically verb-final, and the tense system is invariably based on two verb stems, present and past. Whereas the present stem continues the Old Iranian present, inherited directly from Indo-European, the past stem is based on a participial form of the verb ending in *-ta*. This participle had an active orientation for intransitive verbs, but was originally passive in the transitive paradigm, as in Old Persian *hamiçiyā hagmatā* (rebels (nom.) assembled (nom. m. pl.)) 'the rebels assembled', *ima tya manā kartam* (this what me (gen.) done (nom. nt.

sg.)) 'this is what was done by me'. The subsequent reanalysis of the passive participle as an active verb leads to ergative past tenses, preserved in a number of languages including Kurdish and Pashto, e.g. Kurdish (Kurmanji dialect) *ez ket-im* (I (abs.) fell (1 sg.)) 'I fell', but *min çîrok xwend* (I (obl.) story (abs.) read (3 sg.)) 'I read a story'. The majority of the modern Iranian languages exhibit various stages in the decay of the past tense ergative system into a nominative one, as preserved in the tenses based on the present stem. Modern Persian is typical here of the final stage, with no traces of ergativity except the form of the first person singular pronoun *man* 'I' (< Old Persian genitive *manā*).

Bibliography

The fullest and most detailed general survey available is Rastorgueva (1979–); planned in five volumes, four have appeared so far: 1 (*Drevneiranskie jazyki*) on Old Iranian (1979), 2 (*Sredneiranskie jazyki*) on Middle Iranian (1981), 3 (*Novoiranskie jazyki: zapadnaja gruppa, prikaspijskie jazyki*) on the South-West Iranian and Caspian languages (1982), 4 (*Novoiranskie jazyki; vostočnaja gruppa*) on the East Iranian languages (1987). Spuler (1958) is the only comprehensive handbook in a language other than Russian, although Payne (1981) gives a short survey of linguistic properties of Iranian languages of the USSR. Oranskij (1963) includes annotated specimens of many of the languages and a useful map. Among bibliographical resources, MacKenzie (1969) is a short survey of Iranian studies and full basic bibliography; Oranskij (1975) is a very thorough bibliographical guide to the Iranian languages of the USSR; Redard (1970) is a comprehensive survey of the study of minor Iranian languages, with full bibliography.

References

MacKenzie, D.N. 1969. 'Iranian Languages', in T.A. Sebeok (ed.), *Current Trends in Linguistics*, vol. 5 *Linguistics in South Asia* (Mouton, The Hague), pp. 450–77.
Oranskij, I.M. 1963. *Iranskie jazyki* (Izd-vo Vostočnoj Literatury, Moscow)
—— 1975. *Die neuiranischen Sprachen der Sowjetunion*, 2 vols. (Mouton, The Hague)
Payne, J.R. 1981. 'Iranian languages', in B. Comrie, *The Languages of the Soviet Union* (Cambridge University Press, Cambridge), pp. 158–79
Rastorgueva, V.S. (ed.) 1979–. *Osnovy iranskogo jazykoznanija* (Nauka, Moscow)
Redard, G. 1970. 'Other Iranian Languages', in T.A. Sebeok (ed.), *Current Trends in Linguistics*, vol. 6 *Linguistics in Southwest Asia and North Africa* (Mouton, The Hague), pp. 97–135
Spuler, B. (ed.) 1958. *Handbuch der Orientalistik. Abt. I, Bd. IV, 1, Iranistik* (Brill, Leiden)

25 *Persian*

Gernot L. Windfuhr

1 Historical Background

1.1 Dialectology

Within the Iranian branch of Indo-European, Persian is a member of the West Iranian group, together with the Iranian languages and dialects spoken in Iran and others spoken also outside of Iran, such as Kurdish and Balochi. Within West Iranian, Persian is a member of the South-Western branch, together with other dialects spoken mainly in the southwestern province of Fars, such as Luri and Bakhtiari.

Persian has various dialects. The three major representatives of these are the Persian of Iran in the west, the Persian of Afghanistan now called Dari in the east and the Persian spoken in Soviet Tajikistan in Central Asia in the north-east. Each again has its own dialectal divisions. The number of speakers in each country is approximately: Iran 30 million, Afghanistan five million, USSR 2.2 million.

Iran is a multi-lingual country. While Persian is the official language of Iran, it is the mother tongue of only about 50 per cent of the population. Speakers of non-Persian Iranian dialects constitute some 25 per cent. The remainder speak non-Iranian languages. Besides Arabic, New Aramaic, Armenian, Georgian and Gypsy, Turkic dialects are the most widely spoken, such as Azerbaidjani in the north-west, the archaic Khalaj in the centre of Iran, Turkmenian in the north-east and Qashqa'i in the south-west. Turkic dialects have virtually erased Iranian in northern Afghanistan and Central Asia except for the Tajiki enclave. The Turkisation of much of these areas began before the end of the first millennium AD and does not seem to have halted yet. (Incidentally, those are the same areas where Iranians first took hold on the plateau some 2,000 years earlier.)

1.2 Origins

The evolution of Persian as the culturally dominant language of the eastern Near East, from Iran to Central Asia to northwestern India until recent centuries, began with the political domination of these areas by dynasties

originating in the southwestern province of Iran, Parsa, later Arabicised to Fars: first the Achaemenids (559–331 BC) whose official language was Old Persian; then the Sassanids (c. AD 225–651) whose official language was Middle Persian. Hence, the entire country used to be called '*persē*' by the Ancient Greeks, a practice continued to this day. The more general designation 'Iran(-shahr)' derives from Old Iranian *aryānām* (*khshathra*) '(the realm) of the Aryans'.

The dominance of these two dynasties resulted in Old and Middle Persian-speaking colonies throughout the empire, most importantly for the course of the development of Persian, in the north-east, i.e. what is now Khorasan, northern Afghanistan and Central Asia, as documented by the Middle Persian texts of the Manicheans found in the oasis city of Turfan in Chinese Turkestan (Sinkiang). This led to a certain degree of regionalisation.

1.3 The Formative Period

None of the known Middle Persian dialects is the direct predecessor of New Persian. There are indications that New Persian developed between the seventh to ninth centuries, the period of the Muslim conquest of Iran and later of the high culture of the Arabic-speaking Abbasid court in Baghdad (c. 750–850), to which Iranians contributed so decisively. The first preserved documents come from the eastern regions: three brief inscriptions dating from the middle of the eighth century found in eastern Afghanistan. They were written in Hebrew characters, indicating the early use of the new vernacular by minorities less dominated by the written standards of the time, i.e. Arabic, Middle Persian or local languages such as East Iranian Sogdian.

It was in the north-east, more distant from the caliphate in Baghdad, where Iranian nationalism reasserted itself by the eleventh century. Persian became the universally accepted language first in poetic diction. The major document of this period is the *Shāh-nāmah* 'The Book of Kings', the monumental epic by Firdausi of Tus in Khorasan about the Iranian glory from creation to the Muslim conquest, written in the early eleventh century in an archaising language which used comparatively few Arabic words. It soon became also accepted as the language of official communication and of prose writing vis-à-vis Arabic, the sacred language of the Qur'an and the 'Latin' of the Muslim Near East. For example, the philosopher Ibn-e Sina, Latinised Avicenna, d. 1047, while mostly writing in Arabic, chose to write his *Metaphysics* in Persian for which he created his own Persian terminology.

The 'Persianists' won over the 'Arabists'. Muslim religious propaganda began to contribute considerably to the ever-extending use of Persian through popularising texts such as commentaries on the Qur'an, lives of saints, edificational and moral and religious treatises.

Until the Mongol conquests in the middle of the thirteenth century, the north-east, with cultural centres such as Samarkand, Bukhara, Balkh, Merv, Herat and Nishapur, continued to be the major area of New Persian and its literature. Thereafter, the focus shifted to the west, a major centre being the city of Shiraz in Fars with its most famous poets Sa'di (d. 1292) and Hafiz (d. 1390), from where it shifted to the north, first to Isfahan, the splendid capital of the Safavids (1501–1731), then, from the first half of the nineteenth century, to Tehran, the new capital of the Qajars (1779–1924).

1.4 Standardisation

Persian appears fairly standardised first in early poetic diction, which shows few dialectal variations by the tenth century. (This may be partially due to standardisation by copyists.) Nevertheless, the peculiarities of the eastern poets, especially in their lexicon, led to the compilation of dictionaries explaining those in 'common' Persian, such as the dictionary by Asadi from the middle of the eleventh century.

The formative period for prose writing lasted until the end of the twelfth century. The utilitarian religious texts, just as scientific, historical, geographic, philosophical and mystical writings, naturally paid less attention to high style than to reaching the local public. They retained a considerable degree of local features (in spite of the hands of copyists). Most of the preserved texts originate in the eastern regions, and as such exhibit a fair degree of linguistic homogeneity.

By the thirteenth century, the beginning of Classical Persian, the regionally marked features had largely disappeared in both poetry and prose. This process is concomitant not only with the expansion of Persian, but also with the shift of cultural centres to the west, specifically to Fars. The literary standard was achieved not only through the efforts of poets and writers but, perhaps most importantly, through the efforts of the court chanceries where guides and textbooks on style and rhetoric were compiled from the tenth century.

The dominance of Classical Persian continued to a considerable degree until the beginning of the nineteenth century. At that time new political, economic and cultural conditions, not least under influence from Europe, sponsored gradual simplifications of style. With it came the acceptance in writing of features of the educated spoken language that had developed in the capital Tehran, at first in journalism, then in prose and finally in poetry. Thus emerged contemporary standard Persian. At the same time, Tajikistan under Russian and Soviet rule developed its own literary language which is based on local dialects and written in the Russian alphabet. Iranian Persian ceased to be the accepted standard. It is still the norm in Afghanistan, but decreasingly so as the official language beside East Iranian Pashto.

1.5 Colonial Persian

Persian was cultivated at the courts of the Ottoman rulers, several of whom are known for composing Persian poetry. Literary Ottoman Turkish is a virtual amalgam of Turkish and Persian (with all of the latter's Arabic loan

elements). Similarly, Urdu, '(the language of the) military camp', developed under heavy Persian influence. Persian first entered India with the conquest of north-west India by Ghaznavid armies in the eleventh century. Four centuries later, Persian in its classical form was chosen as the court language of the Mogul kings (1530–1857), who were major patrons of Persian literature and poets from Iran, unlike the contemporary Safavids in Iran. It was at the courts of India and Turkey where many of the major traditional dictionaries of Persian were compiled from the fifteenth to the eighteenth centuries, many with grammatical treatises. Simultaneously, there developed in India a Persian vernacular, and it was from the Indian scribes and secretaries that the English officers of the East India Company, many of whom wrote grammars of Persian, learned their Persian, with all its local idiosyncrasies. Persian was abolished in its last official bastion — the courts of law — in 1837 by the authorities of the East India Company.

2 Phonology

2.1 Sound System

The sound system of contemporary standard Persian is quite symmetric. Its 29 segmental phonemes consist of four pairs of stops and four pairs of fricatives, two nasals and two liquids, three glides, and three pairs of vowels.

Table 25.1: The Persian Phoneme System

Stops	tense/voiceless	p	t	č	k
	lax/voiced	b	d	ǰ	g
Fricatives	tense/voiceless	f	s	š	x
	lax/voiced	v	z	ž	q
Nasals		m	n		
Liquids		l	r		
Glides		y	h	’	
Vowels	tense/long	i	ā	u	
	lax/short	e	a	o	

2.2 Writing System

The Persian writing system uses the Arabic alphabet, which is a consonantal system (see the chapter on Arabic). Vowels are written as follows: long vowels are represented by the letter of the consonant nearest in pronunciation. Thus, the letter ‹y› represents both /y/ and /i/, ‹w› both /v/ and /u/, and ‹alef› both the glottal stop /’/ and /ā/. Short vowels may be, but are usually not, represented by diacritics which ultimately derive from the same letters ‹w›, ‹y›, and ‹alef›. The main innovations in Persian are two: unlike Arabic, short vowels are always represented by consonantal letters in final position, final /o/ by ‹w›, and final /e/ and /a/ by ‹h›. Also, 'Persian'

letters were created for the four Persian consonants /p/, /č/, /g/, /ž/ by adding three dots to the 'Arabic' letters ‹b›, ‹ǰ›, ‹k›, ‹z› (the dots merged into an oblique stroke in the case of ‹g›). The Persian alphabet is given in table 25.2.

Table 25.2: The Persian Alphabet

Alone	*End*	*Middle*	*Initial*			*Name*
ا	ا	ا	ا	’		alef
ب	ب	بـ	بـ	b		be
پ	پ	پـ	پـ	p	P	pe
ت	ت	تـ	تـ	t		te
ث	ث	ثـ	ثـ	s	A	se-ye senokte
ج	ج	جـ	جـ	j		ǰim
چ	چ	چـ	چـ	c	P	če
ح	ح	حـ	حـ	h	A	he-ye ǰimi
خ	خ	خـ	خـ	x		xe
د	د	د	د	d		dāl
ذ	ذ	ذ	ذ	z	A	zāl
ر	ر	ر	ر	r		re
ز	ز	ز	ز	z		ze
ژ	ژ	ژ	ژ	ž	P	že
س	س	سـ	سـ	s		sin
ش	ش	شـ	شـ	š		šin
ص	ص	صـ	صـ	s	A	sād
ض	ض	ضـ	ضـ	z	A	zād
ط	ط	طـ	طـ	t	A	tā
ظ	ظ	ظـ	ظـ	z	A	zā
ع	ع	عـ	عـ	’	A	’eyn
غ	غ	غـ	غـ	q		qeyn
ف	ف	فـ	فـ	f		fe
ق	ق	قـ	قـ	q		qāf
ک	ک	کـ	کـ	k		kāf
گ	گ	گـ	گـ	g	P	gāf
ل	ل	لـ	لـ	l		lām
م	م	مـ	مـ	m		min
ن	ن	نـ	نـ	n		nun
و	و	و	و	v		vāv
ه	ه	هـ	هـ	h		he-ye dočašm
ی	ی	یـ	یـ	y		yā
ء	أ	ئـ	أ	’		hamze

A=letters occurring mostly in Arabic loanwords; P=letters found in Persian only.

The Arabic orthography, the pharyngeal consonants of which are not phonemically distinct in Persian, is retained in all Arabic loans. Other than in Arabic loans, the orthography of Persian is basically phonemic, except for the writing of short vowels discussed above, only rarely using a pharyngeal letter such as ⟨ṣ⟩ in ⟨ṣad⟩ /sad/ 'hundred'.

2.3 Features

In spite of systemic simplicity, there remains considerable debate about the features distinguishing both individual phonemes and sets of phonemes, and about their development. A particularly interesting point is the degree of integration of the foreign loan component, most importantly Arabic, into the system inherited from Middle Persian.

Consonant gemination is a distinctive characteristic of Arabic, whereas in Persian it is a marginal feature. While probably retained in Classical Persian, and still in poetry, it is eliminated in the standard pronunciation of today; for example, Persian *matté* 'drill', Arabic *talaffóz* 'pronunciation' today are pronounced /mate/, /talafoz/.

The highly developed consonantal system of Arabic is considerably reduced in Persian. The non-strident interdental fricatives *θ* and *ð* merged with the respective strident fricatives *s* and *z*. Similarly, the distinctively Arabic pharyngeals merged with non-pharyngeals. Two of the more complex mergers are the following.

The phoneme *q* is intriguing because of its diverse origins and its present articulation and conditioned variation. On the one hand, it originates in an indigenous Persian/Iranian voiced velar fricative with limited functional load. On the other hand, it originates in loans. It represents the merger of the Arabic uvular voiceless stop *q* with the uvular voiced fricative (represented by the respective Arabic letters *qaf* and *ɣeyn*), as well as the voice-neutral back velar stop before back vowels in Turkish (represented by *either* of the Arabic letters). Its peculiar Persian articulation appears like a virtual compromise of its origins: intervocalically it is a voiced fricative; in initial and final position it is partially or fully devoiced, following the devoicing rule, and may have an affricate-like voiced release before vowels (varying with the speaker).

In Persian, glottalic vocalic onset is an automatic feature before initial vowels and in hiatus and as such was originally not phonemic. Arabic, however, has a phonemic voiced pharyngeal ʿ (represented by the letter *'eyn*) and a glottal stop ʾ (represented by ⟨alef⟩ or the diacritic ⟨hamze⟩), which may occur in any position. It is the latter which represents the Persian glottal stop and hiatus in writing, e.g. onset *'in* /ʾin/ 'this', hiatus *pā'íz* /pāʾiz/ 'autumn', affixal hiatus *xāné-i* /xāne-ʾi/ 'a house', *qahve-í* /qahve-ʾí/ 'brown (coffee-ish)'. Phonemically, in Persian the pharyngeal merged with the glottal and with vocalic onset.

2.4 Syllable Structure

The syllable structure of Middle Persian generally reflected that of Old Iranian. This included initial consonantal clusters, which were broken up in Early New Persian by the insertion of a vowel, e.g. MP *brādar* > NP *barādár* 'brother', or by initial vowel, e.g. MP *brū-g* > NP *abrú* 'brow' (so mostly if initial sibilant; note modern loans like *estudiyó* 'studio'). This structure thus agrees with that of the Arabic loan component which has only initial CVC. Since the automatic onset before initial vowel has become phonemic, all Persian syllables now have initial CV, e.g. *in* → /ʾin/ 'this'.

Vowels may be followed by none, one or two consonants, i.e. CV, CVC, CVCC. This makes syllabic boundaries predictable: in any sequence, the consonant immediately preceding a vowel begins a new syllable. This structure has also implications for the status of the two diphthongs of Persian, formerly *ai*, *au*, today assimilated to *ey*, *ow*. Since these are never followed by two consonants like the other vowels, they must be interpreted as a sequence of short vowel + glide, e.g. *dowr* 'turn' as CVCC. They have thus no independent phonemic status, just as in Arabic.

2.5 Stress

The basic stress pattern of Persian is predictable and non-phonemic. Word stress is progressive, i.e. on the last non-enclitic syllable. Phrase stress is regressive. This is evident in pseudo-pairs like *bāz-kón* 'opener' : *bắz kon* 'open!' (*kon* 'to make, do'), where the compound noun has final stress and the verb phrase has stress on the initial member. The third rule, continued from Indo-European, is that stress is on the initial syllable of the vocative noun or phrase, e.g. *xànandé-y-e azìz* → *xắnandè-y-e azìz* 'Dear reader!'

2.6 Morphophonemic Alternation

Unlike Eastern Iranian languages such as Pashto, the rules of morphophonemic alternation of Old Iranian had already ceased to be productive in Persian by the end of the Achaemenid period (c. fourth century BC). This alternation is fossilised in the present and past stems of the so-called irregular verbs and in root nouns. Of course, other changes have long since distorted the regular alternation. Moreover, many such verbs have become regularised and their old past stems lost, a process which has been especially observable in recent centuries.

A considerable portion of the morphophonology of Arabic has been borrowed together with the lexicon. Most complex is that of the verbal system as reflected in verbal nouns and participles borrowed into Persian; to cite only a few frequent forms of the root *n-z-r* 'see, watch': *nazár* 'view', *nazír* 'similar, like', the passive participle *manzúr* 'considered, intended', also 'viewpoint, opinion', the verbal noun of the Arabic eighth formation *èntezắr* 'expectation' with the participle *mòntazér* 'expecting, waiting'.

Probably the most conspicuous part of borrowing is the Arabic plural. Its

complex morphophonology has generally been accepted as an integral part of Persian. The many classes of broken plurals are retained to a considerable degree, varying with the word, certainly with style and possibly with semantic field. The extent of such borrowing has induced the authors of many grammars of Persian to include a considerable section on Arabic morphophonology. However, unlike English which has reanalysed Romance to a certain degree (e.g. 'to **re**-do'), in Persian Arabic morphophonology only applies to Arabic loans and it is not productive, certainly not with the uneducated speaker, rarely affecting Persian words, other than those borrowed early into Arabic and then borrowed back, e.g. *gauhár* > Ar. *ǰauhár* 'essence, jewel', pl. *ǰàvāhír*, and was then borrowed back into Persian.

3 Morphology and Syntax

In terms of morphology Persian with its dialects may be called the most atypical Iranian language. It is to Iranian what English is to Germanic. Unlike East Iranian Pashto and many smaller dialects, it has almost completely lost the inherited synthetic nominal and verbal inflection and their inflectional classes, and thus the *inflectional* distinction of case, number and gender as well as of tense, mood, aspect and verbal gender. This process began already in late Old Persian times. Person and number are, however, distinguished, so is human and non-human gender. The pronouns and endings are shown in the chart given here.

	Singular			*Plural*		
Pronouns	1	2	3	1	2	3
Independent	man	to	u	mā	šomā	išān/ān-hā
Suffixed	-am	-at	-aš	-emān	-etān	-ešān
Endings						
Present stem	-am	-i	-ad	-im	-id	-and
Past stem	-am	-i	-Ø	-im	-id	-and
Perfect stem/'to be'	-am	-i	-ast	-im	-id	-and

The second person singular imperative ending is zero, the second person plural ending is *-id*.

The independent and suffixed pronouns alternate in dependent noun constructions, e.g. *ketāb-e man/ketāb-am* 'my book'. The three sets of personal endings differ only in the third person singular. The third set is in fact the substantive verb 'to be', which is always enclitic, as opposed to the existential *hast-* 'to be (there)', which takes the endings of the past stem.

Pronouns and endings distinguish between human and non-human. All independent pronouns refer to humans only. Thus *u* only means 'he/she', *išān* has become almost exclusively used for third person singular in polite phraseology and has been replaced as a plural by the unmarked *ān-hā*. Non-human items are referred to by the use of the demonstratives *in/ān* 'this/that'. There is no equivalent of 'it' in Persian. This distinction is also found in the interrogative and indefinite pronouns, *ki* 'who' : *če* 'what', *hár-ki* 'whoever' : *hár-če* 'whatever'. Moreover, non-human plurals do not require plural pronouns or endings; their plural marking seems to imply individuation.

3.1 Nouns and Noun Phrases

3.1.1 Nominals

Nouns are simple or compound, based on nominal or verbal stems, e.g. *sāhéb* 'owner', *xāné* 'house', *sāhèb-xāné* 'landlord', *hāvá* 'air' *-peymá* 'to transverse', *-bar* 'to carry', [*havà-peymà*]-*bár* '[aircraft] carrier'; or are nominalised noun and verb phrases, e.g. *ràft-o-āmád* 'traffic', past stems of *raft-án* 'to go' and *āmad-án* 'to come', *bād be-zán* 'fan' lit. 'hit wind'.

There are numerous derivational suffixes. The two semantically least restricted ones, which can be freely added even to phrases are: the abstract suffix *-í*, e.g. *mard-í* 'man-liness', *bozorg-í* 'great-ness', *malèk-o-š-šo'arā-í* 'the status of being poet laureate', and the homophonous denominal relational suffix *-í*, e.g. *ìrān-í* 'Iran-ian', [*zèdd-e irān*]-*í* '[anti-Iran]-ian'.

The comparative suffix is *-tár*, e.g. *bozorg-tár* 'great-er'; the ordinal suffix is *-óm*, e.g. [*paǰàh-o yek*]-*óm* 'fifty-first' (except for Arabic *avvál* 'first' and *āxár* 'last').

3.1.2 Noun Phrases

The basic structure of the noun-adjective phrase and the noun-noun phrase is as follows (N = noun, A = Adjective):

NA: *in* – Measure, Number, Kind–Noun–*há*–*e*–Adjective–*i*
ān
NN: NA^1–*e*–NA^2
NA–Personal Suffixes

The general plural marker is *-há*, and *-án* for adjectival and indefinite pronominal human plurals, e.g. *bozorg-án* 'the elder (people), leaders', *digar-án* 'the others'. The latter is also used for human and human-related plural in literary registers. In addition, there are the plurals of the Arabic loan component which tend to function as a marker of a complex unit. Thus, the plural of *taráf* 'side, direction', *atráf*, has developed the connotation 'surroundings, about', the plural of *vaqt* 'time', *owqát*, generally means 'humour, mood', the loaned feminine-abstract plural *-āt* generalises, e.g. *deh-át* 'the rural area' vis-à-vis the Persian plural *deh-há* 'villages'.

The indefinite marker for both singular and plural is *-i*, e.g. *ketáb-i/ketāb-*

hấ-i 'a book/(certain) books'. It follows the adjective, but often the noun in the presence of more than two adjectives.

Measure, numbers and kind precede the noun and in turn are preceded by the demonstratives *in*/*ān* 'this/that', e.g. *sé* (*tā*) *ketàb* 'three (items) of books', *ín do now' qālì* 'these two kinds of carpet'.

Dependent nominals follow the head noun and are connected by *-e*, e.g. *ketấb-e bozorg-tàr* 'a larger book'. The general function of this construction with dependent nouns and noun phrases, traditionally called *ezāfe* 'addition', is the identification of class and item, the latter ranging from persons, to names and names of species, to numbers, e.g. *ketàb-e mán* 'the book of me/my book'; *xānòm-e J̌avādí* 'Mrs Javadi', *hasàn-e mokrí* 'Hassan Mokri', *gòl-e róz* 'the rose(-flower)', *sā'àt-e sé* 'three o'clock', *dàrs-e haft-óm* 'the seventh lesson'.

3.1.3 Topicalisation

The unmarked sequence head–*e*–dependent is inverted to dependent–Ø–head by topicalisation, most prominently with noun–adjective, noun–comparative, and noun–ordinal, e.g. *kàr* [*-e xúb*]*-i* → [*xúb*] *kàr-i* 'good work', *fìlm* [*-e beh-tár*] → [*beh-tar-ín*] *fìlm* 'the best film' (the so-called superlative), *sāl-gàrd* [*-e sad-óm*] → [*sad-om-ín*] *sāl-gàrd* 'the hundredth anniversary'.

3.2 Single Clauses

Subjects are formally unmarked, indirect objects are in general marked by the preposition *be*, direct objects are marked by the postposition *rā* if specific, adverbial phrases are marked by the prepositions *az* 'from, by, than', *bā* 'with', *tā* 'till, than (comparing clauses)', *dar* 'in/into', *be* 'to' and other functions. The latter two may be elided. These combine with nouns to give numerous adverbial phrases such as *ba-rấ-y-e* 'for the reason of, for', (*be*/*dar*) *rú-y-e* '(to/on) the face of, on, onto' largely supplanting *bar* 'on'.

Persian is an SOV language. The unmarked sequence of the parts of speech in all clauses is subject–adverb–object–verb. Interrogatives do not change this sequence, but occur where the respective answer would be, e.g. (*to*) *ketàb-rā be kí dād-i* lit. 'you the book to whom gave?'. Inversions only occur through topicalisation. In general, sentence-initial and preverbal positions are topical, e.g. *be ù ǰavấb dād-am*/*ǰavàb be ú dād-am* 'I gave him an answer/I gave an answer to him'.

3.3 Categories

In spite of the relative simplicity of the formal aspects of the noun phrase, the syntactic-semantic aspects present problems many of which have not yet been solved. The major ones involved are genericity, definiteness, specificity and reference.

3.3.1 Genericity and Plural

Any unmodified noun in Persian may be generic and imply single or more items, whether subject, predicative complement, direct object or other, e.g. *man ketấb lāzèm dār-am* 'I need a book/books', *ketàb mofíd ast* 'a book is/books are useful', *àn ketấb ast* 'that is a book/those are books' (note the singular pronoun *ān*). This function is exploited in compound verbs (see discussion below), where the verbal content is expressed by a noun followed by a small set of function verbs, e.g. *kàr kard-án* 'work-doing/working', *tarǰomè kard-án* 'translation-making/translating'.

Accordingly, plural is not obligatory when more than one item is implied, unlike English, and plurals in Persian have a more restricted function. The condition for plural marking is restriction of genericity, by reference to specific items or simply by qualifying attributes, as in *u mehmấn dār-ad* 'he has a guest/guests' vs. *u mehmān-hà-y-e āmrikā-í dār-aḍ* 'he has American guests'. This applies, of course, to covert reference as well, as is seen in the pair *ān-hà mo'allém–Ø hast-and* 'they are teachers' vs. *ān-hà mo'allem-há hast-and* 'they are **the** teachers'. This distinction is, however, neutralised after numbers, where plural is never marked.

The basic function of *hā* is not plural, but 'amplification'. While this is interpreted as plural with count nouns, it expresses increase or extent with mass mouns, e.g. *āb-hấ* 'waters, all kinds of waters, plenty of water', and generalisation with adverbs, e.g. *bālā-hấ-y-aš* 'somewhere up there'. This function is most conspicuous with generic objects which remain unmarked, as mentioned. In that case, the presence of *hā* does not express plural, even with count nouns (for specific objects see discussion below), but amplification, e.g. *mà mehmấn–Ø dār-im* 'we have guests' vs. *mà mehmān-hấ dār-im* 'we have lots of, all kinds of guests'.

3.3.2 Genericity and Indefiniteness

Persian distinguishes between genericity and indefiniteness, which latter is marked by the clitic *i*. It occurs with count and mass nouns as well as with singular and plural. As such, it marks restrictive selection out of a generic unit or out of a plurality, e.g. *ketấb-i* 'some/a book' and *ketāb-hấ-i* 'some books', *āb-ǰów-i* 'some, a beer' and *āb-ǰow-hấ-i* 'some kinds of beer'. This function is clearly evident in compound verbs where the presence of *i* eliminates genericity, as in the pair *kấr mi-kon-am* 'I am working' vs. *kấr-i mi-kon-am* 'I am doing something/some work, I am working some/a little'. The restrictive-selective function of *i* is distinct from that of *yek* 'a, one', which counts an item or a group of items. Unlike English 'a' and 'one', both are compatible in Persian, e.g. *yek ketấb-i be-deh* 'give me a (one, some) book'.

There is, however, the similarity between the two languages in that indefiniteness may refer either to specific items known to the speaker or to non-specific items, e.g. *dombàl-e apārtemấn-i mi-gard-am* 'I am looking for

an apartment' may either imply a specific apartment (which I read about in the papers), or any apartment (that will do). In either case indefiniteness is opposed to genericity, as in *dombāl-e apārtemān mi-gard-am* 'I am apartment-hunting'.

3.3.3 *Rā*

Unlike indefiniteness, definiteness is not formally marked in Persian and is only evident in the presence of inherent definites such as demonstratives, personal pronouns, superlatives and ordinal numbers, proper names etc. Thus, the sentence just cited as generic may likewise be interpreted as definite in another context: 'I am looking for **the** apartment'. Until recently it was assumed that there is at least one marker of definiteness, if only with definite direct objects, viz. the postposition *rā*, which was said to be obligatory with such objects. However, not only are there definite direct objects without *rā*, but *rā* is also compatible with indefinite *i*. What is marked by *rā* is not definiteness, but topicalisation or specificity. Thus, since all definite direct objects are normally, but not necessarily specific-referential, they are normally marked by *rā*. It also follows that *rā* is compatible with the indefinite marker *i*, if the latter is specific and implies a unique referent 'a certain, some'. For example, one of the environments where an indefinite is likely to refer to specifics is in sentences with past verbs, as in *xāné-i-rā ātèš zad-and* 'they burned a (certain) house' as opposed to *xāné-i ātèš zad-and* 'they burned a house'. (The sequence indefinite *i* – topicalising *rā* may be roughly compared to the indefinite-specific use of 'this' in colloquial English as in 'they burned **this** house, you know', which refers to a house only known to, or seen by, the speaker.)

While *rā* overwhelmingly topicalises direct objects, it is not confined to them. Thus, it occurs with adverbial phrases of temporal and spatial extension, e.g. *em-šáb-rā in-jấ bāš* 'be/stay here (for) tonight', *hamé-y-e šàhr-rā gàšt* 'he walked all around the city'. Neither with such adverbial phrases nor with direct objects is *rā* obligatory unless topicalisation is involved. This explains why *rā* may be absent in spite of definiteness in sentences like *pà tu káfš kard o ràft* 'she put (her) feet ('foot') in her shoes ('shoe') and left' vs. topicalised *pa-hấ-aš-rā tu kàfš kard o ràft* 'she put **her** feet in her shoes and left' and *èšq né-mi-fahm-ad* 'he does not understand love' vs. *éšq-rā né-mi-fahm-ad* 'he does not understand the notion of love/what love is'.

The topicalising function is also found in highly literary registers, where *rā* may occur in initial phrases, such as [*došmán-rā*] ... *hamé darb-hà-rā be ru-ye ù mí-band-im* 'as to the enemy, we will close all doors except ...' (note the direct object *darb-hā-rā*). The initial phrase *došman-rā* here may well be interpreted as indirect object 'for the enemy'. In fact, there is a small number of verbs where the indirect object is marked by *rā*, such as *má-rā dād* 'he gave (it to) me' side by side *be mán dād*. *Rā* as opposed to *be* appears thus to topicalise these indirect objects as well.

3.3.4 Personal Suffixes

The personal suffixes express not only the experiencing indirect object, but also any direct object: in opposition to topicalised definite direct objects marked by *rā* they express definite non-topical direct objects, e.g. *man ù-rā díd-am* → *díd-am-aš* 'I saw him'. In fact, the independent personal pronouns are always topical. Thus, it follows that independent possession always requires the independent pronoun, e.g. *màl-e mán* 'mine' lit. 'possession of mine'. By contrast, the corresponding suffixes are always non-topical. In addition to the cases mentioned, they function as non-topical objects of prepositions, e.g. *az ù porsíd-am* → *àz-aš porsíd-am* 'I asked (of) him', and as possessors in noun phrases, e.g. *ketàb-e ú* → *ketấb-aš* 'his book'.

In the latter function, they also participate in a remarkable noun phrase inversion, possessor topicalisation: the dependent noun, i.e. the possessor of the subject phrase, is replaced by the respective unstressed suffix, and is itself placed in clause-initial position assuming primary stress so that both bracket the head noun, e.g. *èsm*[*-e ín āqā*] *číst* → [*ín āqā*] *èsm*[*-aš*] *číst* 'what is the name of this gentleman'. With pronouns, there is a threefold gradation: *pedàr*[*-am*] *ostấd ast* → *pedàr*[*-e mán*] *ostấd ast* → [*mán*] *pedàr*[*-àm*] *ostấd ast* 'my father/mý father/me, my father is a professor'.

The [non-topical:topical] function of the pronouns is most widely utilised in the colloquial language where, for example, the indirect construction is expanding. More widely than in the standard language, it functions as the non-topical correlate of direct active constructions, e.g. *gárm* [*hast-*]*am* 'I am warm' → *gárm-am ast* 'I feel warm' lit. 'to me it is warm'. Pragmatically this gives the speaker the option to describe himself as the 'object' of such mental and bodily sensations which are 'coming or happening to him' without his doing, or as the 'subject' with his active involvement.

Similarly, the possessive construction with *dāšt-án* 'to have' may alternate in colloquial speech with the suffixal construction, as long as no true possession is implied, e.g. 'he is two years old' may be expressed as *ù dó sāl dār-ad* 'he has two years' or as *dó sāl-eš e* (← *ast*) 'two years are to him'.

It is evident, then, that the personal suffixes have the general function of what may be called non-topical 'oblique case'.

3.4 The Verb Phrase

The basic verb system of contemporary Persian may be as given in the chart using the verb *rav/raft* 'go' in the third person singular with negation. As is evident, several of these verb forms have double function.

	Indicative	*Non-Indicative*	
Imperfective:			
Present	né-mi-rav-ad	bé-rav-ad/ná-rav-ad	Subjunctive
Past	né-mi-raft	né-mi-raft	Counterfactual
Inferential Past	né-mi-raft-e ast	né-mi-raft-e ast	Counterfactual
Aorist:	ná-raft	ná-raft	Subjunctive

Perfective:			
Present	ná-raft-e ast	ná-raft-e bāš-ad	Subjunctive
Past	ná-raft-e bud	ná-raft-e bud	Counterfactual
Inferential Past	ná-raft-e bud-e ast	ná-raft-e bud-e ast	Counterfactual

The stative verb *bud-án* 'to be' has only an imperfective subjunctive without *be-*, *bāš-ad*, and no past perfect, but a literary present *mi-bāš-ad*. *Dāšt-án* 'to hold, keep, have' has only a perfective subjunctive, *dāšt-é bāš-ad*. Neither has *mi-* when used as imperfective past and counterfactual. This restriction does not apply to the use of *dāšt-án* in compound verbs.

The verb forms are based on three stems: present, aorist and perfect, the last regularly derived from the aorist stem by *-e*. All perfect forms are periphrastic with forms of the verb 'to be'. The imperfective prefix *mi-* occurs with all three stems, while the subjunctive prefix *be-* occurs only with the present stem and is mutually exclusive with negation.

The nominal forms are the three stems and the verbal noun, called 'infinitive', marked by *-an* as in *raft-án* 'to go, going'.

3.4.1 Categories

This verb system used to present considerable problems. Until very recently a good many grammars and textbooks omitted some of the more complex forms, while others postulated non-existing, usually obsolete, forms. And if the complex forms were mentioned, their function was mostly only circumscribed.

3.4.2 Aspect and Tense

The key to the understanding of the system is the recognition of the functions of the forms marked by *mi-*, of the forms marked by the perfect stem in *-e* and, most importantly, of the aorist *raft* which used to be identified as (simple) past or preterit for the obvious reason that this is the general form used in simple past narrative. With the 'past' *raft* opposed to the present *mí-rav-ad*, there appeared to be a system based on tense distinction, quite similar to Western European systems, notably the French system as traditionally understood. This was reinforced by the pair of the present and past perfects *raft-é ast* and *raft-é bud* and the imperfect *mí-raft*.

However, aspect is as basic a categorical vector of the system as is tense. *Mi-* is the marker of imperfectivity. As such it may express habitual action, progressive-ingressive action, as well as future action in the present and past, e.g. present *hamišè/al'ān/fardà kár mi-kon-am* 'I always work/I am working (right) now/I will work, will be working tomorrow', past *hamišè/dirùz/fārdà kár mi-kard* 'he was always working, would always work/he was working yesterday (when I came)/(he thought:) he would work, would be working the next day', the latter in contexts such as anticipation in an interior monologue.

The perfect forms are not simply perfective, but resultative-stative. This is most evident with change-of-state verbs, e.g. *hasán ān-jà nešast-è ast/bud* 'Hasan has/had sat down there' = 'Hasan is/was sitting there', *Maryàm lebàs-e qašáng-i pušid-è ast/bud* 'Maryam has/had put on a nice dress' = 'Maryam is/was wearing a nice dress'. Both occur also in a future context, e.g. *fardà sā'at-e sè raft-é am/raft-é bud-am* 'by three o'clock tomorrow I will be gone/by three o'clock the next day I would be gone', the latter again in anticipation in the past.

Most instructively, the aorist is not confined to past contexts, but occurs in present and future contexts as well, most evident with verbs implying motion, e.g. in a past context *hasán diruz be bāzár raft va ín-rā xarid* 'Hasan went to the market yesterday and bought this', in a present context *to bāš-i, man ráft-am* 'you stay here, I am on my way/am going now', which may be said when still seated, or in a future context *šàyad má ham raft-im* 'we will most likely go, too', said after hearing that someone will go to see an exhibition. The future use of this form is largely confined to the colloquial language. In educated registers a formation with *xāh*, the unmarked present stem of *xāst-án* 'to want, will', is used followed by the uninflected form, *ná-xāh-ad raft* 'he will not go'.

The aorist does thus certainly not indicate past tense; rather, it is tense-neutral and it is the context which identifies time. It is a member of both the present and past subsystems, and therefore is called here 'aorist'.

3.4.3 Inferential Past

The complex forms *mí-raft-e ast*, which combines imperfective *mi-* with the perfect *-e*, and *raft-é bud-e ast*, a double perfect, express remote past in the literary register. However, they are not confined to literary style, but are as frequent in the colloquial language without referring to remote past. What they express is the category of inference, that is mainly second-hand knowledge, conclusion and reminiscence. In this they are joined by the perfect form *raft-é ast* which also functions as the inferential aorist. All three forms of the inferential past are thus derived from the perfect as is the case in a good number of other languages which have that category. To give one example: *zāher-án nevisandé, vàqt-i ān nāmè-rā mi-nevešt-é* (*ast*), *xód-aš-rā bā ín āmpúl-i, ke ruz-e qàbl xarid-è bud-é* (*ast*), *košt-é* (*ast*) 'apparently, the writer killed (*košt-e ast*) himself with this injection, which he had bought (*xarid-é bud-e ast*) the day before, while he was writing (*mí-nevešt-e ast*) that letter'. The non-inferential past forms in this context would imply a fact or be at least uncommitted.

The tense opposition [present:[past:inferential past]] is therefore likewise a fundamental vector of the system. Future, however, is not a tense, but at best a modality. As is evident in the examples above, all present and past forms may be used in a future context.

3.4.4 Mood

The basic function of the subjunctive is to express potential action. As such it functions as adhortative, e.g. *bé-rav-ad* 'he should go/let him go'. It is obligatory after verbs with potential connotations such as modal verbs and expressions and verbs like 'to fear/be afraid to', 'to hope to' etc., e.g. *bà-y-ad bé-rav-ad* 'he must go', *mì-tars-ad bé-rav-ad* 'he is afraid to go'. (The infinitive-verbal noun is strictly nominal and expresses 'the going' rather than 'to go'.)

The basic function of the counterfactual is to express actions or states which are unlikely to, or did not, come about. As such it functions in wishes and hypothetical statements. It is thus tense-neutral, and the distinction is strictly one of aspect, e.g. *kàš mí-raft* may be interpreted as 'if he would only go' or 'if he had only gone'. Similarly, the perfective, e.g. *kàš raft-é bud* is either 'if he were only gone' or 'if he had only left'.

In connection with necessity, it also expresses an action which should have, but did not, happen, as well as an action which had to be done instead of another, e.g. *bà-y-ad fardá mi-resid* 'he should arrive, have arrived tomorrow (but now they say...)', *tázè qàbl-aš ham bà-y-ad mí-raft-im qazá be-xor-im* 'we first had to go to have some food (and thus did not come)'.

3.4.5 Causation

The causal suffix is *ān*, e.g. *xor* 'to eat' vs. *xor-án* 'to make eat, feed', *rav* 'to go, leave' vs. *rān* 'to drive' (< *rav-ān*). Today, this suffix appears to be increasing in productivity, perhaps due to increased linguistic consciousness of writers. But it had been on the decline along with the general tendency, beginning in Early New Persian, to replace simple verbs by compound verb constructions consisting of a nominal followed by a relatively small set of verbs, the most frequent of which are *kard-án* 'to do, make' and *šod-án* 'to become' (originally 'to go'). These two function as markers of causality. Three stages of causation are distinguished: in simple inherently causative verbs, agent mentioned is expressed actively, agent implied by the third person plural ending, agent not implied by the perfect participle + *šod-án*, e.g. *dár-rā bàst* 'he closed the door', *dár-rā bàst-and* 'they/someone closed the door', *dàr bast-é šod* 'the door closed/was closed'. In compound verbs, *kard-án* assumes the causative function, e.g. *ù-rā bidár kard* 'he woke him up', *ù-rā bidár kard-and* 'they/someone woke him up', *bidár šod* 'he woke up'.

The non-agentive construction with *šod-án* has generally been identified as passive, since with inherently causative verbs it appears like a Western European passive, e.g. *košt-é šod* 'he got killed' is assumed to be a equivalent to 'he was killed'. The Persian passive, however, is strictly agentless: unlike English (*he was killed **by X***), it excludes the expression of a known agent. Moreover, it is confined to causal verbs, which may imply a change of state, such as *košt-án* 'to kill', creation, such as *nevešt-án* 'to write', *sāxt-án* 'to build', movement of an object, such as *āvard-án* 'to bring', and observation, such as *nešān dād-án* 'to show'. Its function as a non-agentive construction is utilised pragmatically whenever the speaker wishes not to mention the agent, as is often the case in bureaucratic jargon and in polite phraseology so typical for Persian.

3.5 Subordinate Clauses

3.5.1 Relative Clauses

Relative clauses are introduced by the general relative pronoun *ke* 'that'. The head noun is taken up again in the relative clause by the respective independent or suffixed pronoun, e.g. *àn márd ke māšín-rā* [*az u*] *xaríd-i* 'that man, from whom you bought the car'. This pronoun is optional if *ke* functions as the subject or direct object of the relative clause.

Restrictive relative clauses are marked by *-i*, e.g. *àn márd-i ke māšín-rā az-aš xarid-í* 'that man from whom you bought the car' (not the other one etc.). This *-i* merges with the homophonous indefinite *-i*, e.g. *márd-i ke zan ná-dār-ad tanhá ast* 'a man who has no wife is lonely'.

3.5.2 Sequence of Clauses

The basic rule for the sequence of main and subordinate clauses in contemporary Persian may be stated as follows: subordinate clauses with actions or states which logically or temporally precede others, i.e. cause, time and condition, precede the main clause; those whose actions and states logically or temporally follow others, i.e. explanation, sudden interruption, time of potential or factual completion and exception, follow the main clause.

This basic rule is seen in the pattern of the most frequent adverbial clauses.

Preceding			*Following*		
Cause	čun	'because'	Explanation	zí-rā	'(that is) because'
Time	váqt-i	'when'	Interruption	ke	'when (suddenly)'
Point/ Stretch	tā	'as soon as' 'as long as'	End point	tā	'until, so that'
Condition	ág‘ár	'if'	Exception	mág‘ár	'unless, if not'

The semantically neutral enclitic conjunction *ke* may be substituted for the conjunctions of preceding clauses, e.g. *čun/váxt-i/tā/ág‘ár pul nà-dār-ám, né-mi-rav-am* 'because/when/as long as/if I have no money I will not go', all → *púl-ke nà-dār-am, né-mi-rav-am*. In addition to these, there are numerous adverbial conjunctival phrases either with nouns, such as (*dar*) *mowqé-i ke āmád* '(at) the moment (that) he came', or with adverbs, such as *pìš az ín ke be-rav-ád* 'before (this that) he left'. Their general structure shows that syntactically they are relative clauses, restrictive relative clauses

with nouns, [N-*i ke*], and non-restrictive with adverbs, [- *in ke*]. Since adverbs are strictly prenominal they require a 'dummy' noun to introduce the dependent clause, either *in* 'this' or less frequently *ān* 'that'.

Object, subject and complement clauses, which express facts or possibilities depending on the main clause, follow the main clause, e.g. object *díd-am (ke) ān-ǰā níst* 'I saw that he is not there', subject *ma'lúm ast ke u níst* 'it is obvious that he is not here', complement *hàqq-aš ín ast ke pùl ná-dār-am* 'the truth of it is (this) that I have no money'. As is evident, the conjunction *ke* is optional with object clauses, but obligatory with subject and complement clauses.

Syntactically, these clauses are relative clauses as well, as seen most clearly by topicalising inversion: *ín ke u níst ma'lúm ast* '(this) that he is not here is obvious', *ín ke u ān-ǰā níst díd-am* '(this) that he was not there I noticed'.

3.5.3 Verbal Categories

The 'logic' of the sequence of clauses is paralleled by the 'logic' of the verbal categories. All subordinate clauses, including relative clauses, strictly follow the semantics of tense, aspect and mood.

Factual actions and states are in the indicative, even in conditional clauses, e.g. [*àgar mí-xāh-i*], *mí-rav-im* 'if you (really) want to, we will go'. Potential actions and states are in the subjunctive in clauses with potential connotation such as final, concessive and conditional clauses, as well as in temporal and relative clauses with implicit condition, therefore also including those with conjunctions like 'before', 'without', e.g. *ráft* [*tā az ù bé-pors-ad*] 'he went in order to ask him', [*àgar/vàqt-i be-rav-ád*] *kàs-i digàr níst* 'if/when he goes there will be no one left', *fárš-i* [*ke gere-hà-y-aš riz-tàr bāš-ád*] *beh-tár ast* 'a carpet the knots of which are finer is better', [*pìš az ín ke bè-rav-í*] *telefón kon* 'before you go, call'. Unlikely or impossible actions or states are in the counterfactual.

Similarly, aspect. Incomplete actions are expressed by the imperfective, resulting states by the stative and completed perfective actions by the aorist. This is true for both the indicative and the non-indicative. Most instructive in this context is the use of the aorist in explicitly or implictly conditional contexts. There it expresses the potential completion as a condition for another action, in contrast with the imperfective subjunctive, e.g. subjunctive [*àgar hasàn be-rav-ád*] *be màn telefón kon* 'if Hassan leaves/should he leave, give me a call', aorist: [*àgar hasàn-rā did-í*] *be màn telefón kon*, [*àgar na-búd*] *yād-dášt-i bè-nevis* 'if/as soon as you find Hassan, give me a call; if he is not there, write a note'.

Finally, tense. Most instructive in this context are object clauses expressing observed facts, including reported speech. Not only do these require the indicative, but also the imperfective or stative present if the action or state is simultaneous with the time of the main verb (whereas in English the tense of the main verb has to be 'mapped' onto the dependent verb), e.g. *vàqt-i resid-ím šeníd-im* [*ān-ǰà čand ruz-è bārán mi-ā-y-ad*] 'when we arrived we heard that it had been raining there for several days', *gòft* [*ke né-mi-ā-y-ad*] 'he said he would not come'. On the other hand, completed past action is obligatorily expressed by the past perfective, e.g. *fàsl-i* [*ke ferestād-è bud-íd*] *resíd* 'the chapter you sent has just arrived' (note the simple past in English).

3.6 Continuity and Innovation

The following is a brief summary of the diachronic development of the forms and categories of Persian and of the main divergences between the three main dialects of Persian. Both reflect the continuity of earlier categorical distinctions as well as the process of ever-increasing differentiation after the collapse of the Old Iranian inflectional system.

3.6.1 Gender

The Old Iranian distinction between masculine, feminine and neuter gender had been lost in late Old Persian. Subsequent stages developed various means of distinguishing between animate and inanimate, as in the case of contemporary Persian, described above.

3.6.2 Noun Phrase

Categories. The history of noun phrase morphosyntax is the history of the foregrounding of genericity, indefiniteness and specificity. Already in Old Persian, the singular could be used generically. However, it was restricted to non-human. This still held in Early New Persian where human plural was marked in predicative position, e.g. *havā-šenās*[*-ān*] *bud-and* 'they were meteorologists'. In contemporary Persian, genericity is generalised.

The indefinite marker *-i* originates in the Old Iranian prenominal number *aiwa* 'one'. In Middle Persian it developed the secondary function of indefiniteness if following the noun. In Early New Persian this use was generalised to singular and plural nouns, but it was still immediately attached to the noun. Today, it generally follows the adjective with a few marked exceptions.

The history of *rā* and of the pronominal suffixes is the coming into syntactic-semantic prominence of the direct object and specificity. *Rā* originates in the Old Persian postposition *rādi* 'by reason of, concerning', cf. Latin *ratiōne*. Thus in Middle Persian *rā* expressed cause, purpose and reference (partially like English '(as) for'). By extension of the implicit directional meaning its range began to include occasional use with indirect and direct objects in Late Middle Persian, a range continued in Early New Persian.

In Early New Persian, *rā* had a similar range, but was not obligatory with either direct or indirect objects. The reduction of its range towards specificity may be shown with the following examples. *rā* marked indirect objects which could be: (a) the beneficiary of an action, alternating with the preposition *ba* 'to'; (b) the possessor, alternating with the verb *dāšt-an* 'to have'; and (c) the experiencer in indirect constructions expressing mental and bodily sensations such as hunger and liking, alternating with the personal suffixes. In contemporary Persian, a virtual semantic-syntactic split has occurred. The three indirect objects are now distinctively marked by the alternates, e.g. *man ō-rā mē-gōy-am* > *man be u mi-gu-y-am* 'I am telling him', *ō-rā du pisar bud-and* 'to him were two sons' > *u do pesar dār-ad* 'he had two sons', *az an ma-rā xwaš āmad* > *az ān xoš-am āmad* 'I liked it'. In the Persian of today, for most other uses *rā* has been preserved in, and was replaced by, the prepositional phrase *ba-rā-y-e X* 'for X'.

Nominal Subordination. The function of nominal subordination to express class-item, among which possession is only one, continues an Old Iranian formation, verbless appositional phrases introduced by the generalised relative pronoun Old Persian *haya*/Avestan *yat* > *-e*. This progressive subordination, NN^1-*e* NN^2, is typically South-Western Iranian in terms of dialectology. The marked topical inversions in Persian are the unmarked ones in North-Western Iranian, and can in part be understood as originally marked borrowed features.

The range of the general conjunction *ke* is the result of the merger in New Persian of three Middle Persian conjunctions, *kē* 'who, which', *kā* 'when' and *kū* 'where'. The use of *-i* to introduce restrictive relative clauses, and thus the marking of restrictiveness of relative clauses in contemporary Persian, is the result of a similar generalisation. It originates in the indefinite marker *-ē*, and was exclusively used in Early New Persian with indefinite head nouns.

3.6.3 Verb Phrase

The endings of the aorist continue the Middle Persian substantive verb 'to be', thus MP *h-am* > NP *-am*. The infinitive-verbal noun continues the Old Iranian verbal noun marked by *-tan-*. The endings of the present continue Old Iranian, and ultimately Indo-European endings, as is evident in the endings of the third persons *-ad* < *-a-t-i*, *-and* < *-a-nt-i*, as is the case with the endingless imperative of the second person singular and the initial stress in the imperative and the vocative.

Aspect. The functions of the three stems of the verb reflect their history. Present stems originate in the Old Persian 'present', i.e. imperfective, stems (e.g. OP *bar-a-* > NP *bar-* 'to carry, bear', *da-dā* > NP *dah-* 'to give', *kr-nu-* > NP *kon* 'to do, make'). 'Past', i.e. aorist, stems originate in the Old

Persian perfect participle in *-ta* (e.g. OP *bṛ-tá* > NP *bord*, *dā-tá* > NP*dād*, *kṛ-tá* > NP *kard*). Functionally, constructions with this participle and the copula served as the successor of the older inflectional forms of the Old Iranian 'perfect' and 'aorist' systems, a process that had begun already in Old Iranian. This construction lost its 'perfect' function in Middle Iranian, and a new perfect stem developed in New Persian and a regionally confined number of other dialects, which is derived from the aorist stem by the substantive suffix *-e* (< *-ag* < *-ak-a*).

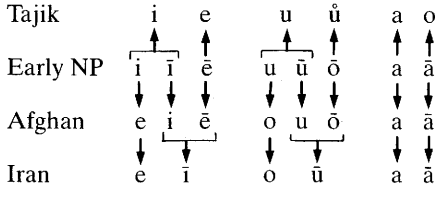
Similarly, the history of *mi-* reflects the evolution of aspect. *Mi-* originates in the Old Iranian adverb *hama-aiwa-da* 'at the same time, place'. Middle Persian *hamē(w)* 'always, continuously', besides its adverbial function, was also used to express durative action or state, which was extended to iterative and distributive function in Early New Persian.

At that stage, habitual action in past and present, as well as counterfactual action, were expressed by *-ē(d)*, which originates in the generalised third person singular optative *hait* 'may it be' in Old Iranian, where optatives had already a secondary habitual past function. This clitic was virtually lost in Classical Persian, and both habitual and counterfactual functions were taken over by *mē-*, by then strictly an aspectual prefix, with the secondary function of counterfactuality together with the past perfect, as is the case in contemporary Persian.

4 Dialectology

The three main dialects of Persian in Iran, Afghanistan and Tajikistan have diverged in their phonology, most prominently in their vocalic systems. The developments in their morphosyntax is the history of the increasing differentiation prominently in their verb systems by the development of new formations expressing aktionsarten, mood and causation, partially under the influence of Turkic.

The development of the vowels is shown in the diagram given here.



Compared with Early New Persian, Afghan Persian is the least changed, lowering the short high vowels as in Iran to mid vowels, which are now opposed to the retained long mid vowels, while the old long high vowels lose their length distinction. Tajiki is the most changed, losing the length

distinction, most likely under the influence of Turkic, by the merger of the short and long high vowels and the rounding of long *a*.

In terms of nominal syntax, the marked inversion of possessor head noun, *pedar-e man* > [*man*] *pedar* [*-am*] 'my father', has become the unmarked construction in Tajiki, again under the influence of Turkic. The colloquial language in Iran has developed a focalising suffix *-é*, e.g. *sag-é* 'the dog mentioned'.

Inference is found in both Afghan and Tajik Persian. Similar forms are found in Early New Persian prose texts, most of which originate in the east, as mentioned, but they disappeared as regionally marked features in Classical Persian. Their appearance in early texts, as well as their reappearance in contemporary standard Persian of Iran, can again be explained by interference from Turkic where inference is marked by *emiš* (see the discussion of Turkish *-mIş*, page 632). Unlike Turkic, inference is not tense-neutral in Persian, but confined to the past. In Tajiki, however, *mi-raft-e ast* has already become tense-neutral.

The verb forms of Turkic are mostly based on participles. In Tajiki, this has resulted in the development of participial formations with so-called con-verbs, where the participial main verb is followed by a varied set of verbs whose meaning is generalised to express various aktionsarten. For example, *šud-an* 'to become' expresses completion, *bar-omad-an* 'to come out of' thorough completion, and *guzašt-an* 'to pass through, by' completion after a prolonged action, as in [*kitob-ro xond-a*] *šud/bar-omad/guzašt* 'he completed reading the book/he completed reading through the book/he completed the book after prolonged reading'.

Similarly, in Tajiki the progressive is a participial formation with *istod-an* 'to stand', as in [*kitob-ro xond-a*] *istod-a ast* 'he is reading the book'. This development has progressed less in Afghan Persian, which has developed two participial formations, the progressive marked by the con-verb *raft-an* 'to go', as in [*ketāb-ra xānd-a*] *mē-rav-ad* 'he is reading the book', and the dubitative based on the particle *xāt* < *xāh-ad* 'it will/may (be)', as in [*zad-a*] *xat bud-om* 'I might hit'.

In contrast, in the formations developing in Iranian colloquial Persian both verbs are inflected as seen in the progressive based on *dāšt-an* 'to keep, hold, have', as in *dār-ad* [*ketāb-rā mi-xān-ad*] 'he is reading/is about to read the book', in the potential progressive in Tehrani based on *raft-an* 'to go' + subjunctive, as in *mi-rav-ad* [*be-suz-ad*] '(the motor) is about to burn', or in the formation expressing sudden action based on *zad-an* 'to hit', as in *zad-and* [*raft-and*] 'off they went'. Similarly, a new causative formation, 'have-other-do', based on *dād-an* 'to give', inflects both causer and caused, as in *raft va dād* [*šāx-hā-y-aš-rā tiz kard-and*] '(the goat) went and had her horns sharpened' lit. 'she gave, they sharpened'.

Participial formations are already found in the early prose texts, most of which originate in the east. For example, continuity was expressed by *dāšt-an* 'to keep, hold, behold' with transitives and by *mānd-an* 'to remain, stay' with intransitives, as in [*girift-a*] *dār-ad* 'he keeps [holding]' and [*halāk šud-a*] *bi-mān-and* 'they will keep [perishing]'. Again, in Classical Persian these eastern features were eliminated.

However, the 'passive' in contemporary Persian does originate in such a formation. In Early New Persian there existed a participial formation based on either *āmad-an* 'to come' or *šud-an* 'to become', earlier 'to go', which occurred with both transitives and intransitives, e.g. [(*ān-rā*) *yād kard-a*] *āmad-a/šud-a ast* 'it has been recalled', and [*būd-a*] *šud/āmad* 'it came into [being]'. In Classical Persian, the use with intransitives and 'come' is lost, and the active participle eliminated: (*ān*) *yād šod-a ast*.

Bibliography

Windfuhr (1979) is the 'state-of-the-art' concise survey of the study of Persian grammar, theoretical approaches and analyses, including new insights into syntax-semantics and phonology with extensive references, together with the most comprehensive alphabetical and topical bibliographies to date. Lazard (1957) is an excellent detailed descriptive-structuralist grammar of contemporary Persian. Phillott (1919) is the most extensively documented grammar of Persian to date, with notes on dialectal variations and many illuminating insights into the pragmatic use of the language. Lumsden (1810) is still the only grammar to make thorough use of the indigenous Muslim grammatical theory and insights, many of which were only rediscovered more recently. Jensen (1931) is a comprehensive descriptive and comparative grammar of Classical Persian with notes on contemporary Persian. For the earlier history, Lazard (1963) provides an abundantly documented analytic description of the Persian of prose texts of the eleventh and twelfth centuries, with historical and dialectal annotation.

Acknowledgement

Work on this chapter was supported by a generous grant from the National Endowment for the Humanities for the comprehensive study of the languages and dialects of Iran.

References

Jensen, H. 1931. *Neupersische Grammatik, mit Berücksichtigung der historischen Entwicklung* (Carl Winter Universitätsverlag, Heidelberg)

Lazard, G. 1957. *Grammaire du persan contemporain* (Klincksieck, Paris)

——— 1963. *La langue des plus anciens monuments de la prose persane* (Klincksieck, Paris)

Lumsden, M. 1810. *A Grammar of the Persian Language; Comprising a Portion of the Elements of Arabic Inflection, Together with Some Observations on the Structure of Either Language Considered with Reference to the Principles of General Grammar*, 2 vols. (Calcutta)

Phillott, D.C. 1919. *Higher Persian Grammar for the Use of the Calcutta University, Showing Differences Between Afghan and Modern Persian with Notes on Rhetoric* (The University Press, Calcutta)
Windfuhr, G.L. 1979. *Persian Grammar. History and State of its Study* (Mouton, The Hague, Paris and New York)

26 Pashto

D.N. MacKenzie

1 Introduction

Long recognised as the most important language of the North-West Frontier Province of British India, now Pakistan, where it is spoken by 90 per cent of the population, Pashto was by royal decree of 1936 also declared to be the national language of Afghanistan in place of 'Dari' Persian. This official preeminence was artificial, however, and it now shares the honour with Persian. The areas of Afghanistan to which Pashto is native are those in the east, south and south-west, bordering on Pakistan, but in recent years Pashto speakers have also settled in parts of the northern and eastern provinces of the country. Reliable census figures of the number of speakers are only available from Pakistan. There, in the fifties, the total number of Pashto speakers was stated to be nearly 5.35 million, of whom 4.84 million (4.47 million of them in the North-West Frontier Province and 270,000 in Baluchistan) claimed it as their mother tongue. In Afghanistan in the same period semi-official estimates gave the number of speakers (presumably including those for whom it was a second language) as between 50 and 60 per cent of the total population of 13 million, i.e. between 6.5 and 7.8 million. Even allowing for some nationalistically inspired exaggeration in these figures, it seems permissible to assume that today at the very least 10 million people in Afghanistan and Pakistan are native speakers of Pashto. In terms of numbers it is, therefore, the second most important of modern Iranian languages.

The name of the language, properly *Paxto*, also denotes the strong code of customs, morals and manners of the Pashtun (*Paxtun*, Indianised as *Paṭhān*) nation, also called *Paxtunwālay* — whence the saying *Paxtun haya nə day če Paxto wāyi lekin haya če Paxto lari* 'A Pashtun is not he who speaks Pashto, but he who **has** Pashto.'

2 History

Pashto belongs to the North-Eastern group within the Iranian branch of Indo-European. The relationship can best be demonstrated by two phonological features characteristic of most members of this branch, viz. the

development of the Old Iranian initial voiced plosives *b*, *d*, *g* and of the dental groups *-ft-*, *-xt-*. Initial *b*, *d*, *g*, preserved in Western Iranian, regularly became the voiced fricatives *β*, *γ*, *δ* in Khwarezmian and Sogdian. For example, Old Iranian *brātar-* 'brother', **buza-* 'goat', **duγdar-* 'daughter', *dasa-* 'ten', *gauša-* 'ear', **gari-* 'mountain' yield Sogdian *βr't*, *'βz-*, *δwγt'*, *δs'*, *γwš*, *γr-*, Khwarezmian *βr'd*, *'βz*, *δγd*, *δs*, *γwx*, *γryck*. Pashto shows the same development of *g-*, in *γwağ* 'ear', *γar* 'mountain'; *b-*, however, has passed through *β-* to the labial continuant *w-*, *wror* 'brother', *wəz* 'goat', and *d-* through *δ-* to *l-*, *lur* 'daughter', *las* 'ten'.

The dental group *-ft-*, also preserved in Western Iranian, becomes voiced in Eastern Iranian to [-βd-]: e.g. Old Iranian **hafta-* 'seven', **tafta-* 'heated', **xšwifta-* 'milk' give Sogdian *'βt*, *tβt*, *xšyβt*, Khwarezmian *'βd*, —, *xwβcy* [*xuβji]. In Pashto the group has been simplified either to *-(w)d-* (cf. Khotanese Saka: *hauda*, *ttauda*, *svīda*), as in *tod*, feminine *tawda* 'hot', *šodə/e* 'milk', or to *-w-*, as in *owá* 'seven'. *-xt-* coincides with *-γd-* in Eastern Iranian, e.g. *suxta-* 'burnt', *baxta-* 'shared', *duγdar-* 'daughter' give Sogdian *swγt*, *βγt-*, *δwγt'*, Khwarezmian —, *βγd*, *δγd*. Just as *-γd-* was reduced in Khotanese, via [-d-], to a hiatus-filling [-w-] (*sūta* [*sūda-] > *-suva*, *būta* [*būda] > *būva*, *dūta* [*dūda] > *dūva*), so in Pashto it has either become *w* or, finally, dropped without trace: *sə́way* 'burnt', *su*, feminine *swa* 'it burnt', *tə* 'went' < **taxta-*, *tar-lə́* 'father's brother's daughter' < **-duγda-*.

The change of *d* to *l*, already mentioned, is found in other neighbouring languages: there is evidence for it having occurred in at least some Sogdian dialects and in Bactrian (e.g. *Βαγολαγγο* < **bagadānaka-*, the modern Baghlan), and it is normal in modern Munji (where *luγda* 'daughter', *pāla* 'foot' < **pādā-*). Pashto goes further, however, in that all dentals, *t*, *θ*, *d*, become *-l-* post- or intervocalically; e.g. OIran. *pitar-* 'father', *sata-* 'hundred', *paθana-* 'broad', **čaθwar-* 'four', **gada-* 'robber', **wadi-* 'stream', yield Pashto *plār*, *səl*, *plən*, *calor*, *γal*, *wāla*. In other contexts though the dentals were often preserved, e.g. *tə* 'thou' < *tú*, *dre* 'three' < **θrayah*, *atə́* 'eight' < *ašta*, (*yaw-*, etc.)*wišt* 'twenty(-one, etc.)' < **wīsati* (contrast *šəl* 'twenty' alone < **wīsáti*).

Only a few other sound changes can be mentioned. Perhaps the most striking in Pashto, as in the Pamir languages, are those undergone by some *r*-groups. Both *-rt-* and *-rd-* changed into the retroflex *-ṙ-*, and *-rn-* into its nasalised counterpart *-ṇ̂-*: e.g. **ārta-* 'milled' > *oṙə́* 'flour', *mṛta-* 'dead' > *məṙ*, **zṛdya-* 'heart' > *zṙə*, **amarnā-* > *maṇ̂á* 'apple', **karna-* 'deaf' > *kuṇ̂*. The presence of a sibilant complicated matters. *sr* and *rš* became *x̌* and *ğ* respectively (on the phonemes written *x̌*, *ğ*, see below), e.g. **hwasrū-* 'mother-in-law' > *xwáx̌e*, **ṛša-* 'bear' > *yağ*, and in *-str-*, *-štr-*, *-ršt-* the *-t-* was lost, leaving *-x̌-*, e.g. *uštra-* 'camel' > *ux̌*, *wāstra-* 'grass' > *wāx̌ə́*, **hṛštaka-* 'left' > *íx̌ay*. *-rs-*, on the other hand, coincided with *-rst-* to yield *-x̌t-*, and *-rz-* similarly gave *-ğd-*, e.g. **uz-kṛstaka-* 'cut out' > *skáx̌tay*, *pṛsa-* 'ask' > *pux̌t-*, **warsya-* 'hair' > *wex̌tə́*, **bṛz-* > *uğd* 'long', **arzana-* 'millet' > *ğdən*. It is an example of this development of *-rs-* that has given *Paxto* its name, from an original **Parsawā-* closely akin to the old names of the Persians and Parthians, respectively *Pársa-* (< **Parswa-* ?) and *Parθawa-*. *Paxtun* probably continues an old **Parswāna-*.

The Pashto lexicon is as fascinating as an archaeological museum. It contains side-by-side words going back to the dawn of Iranian, neologisms of all ages and loanwords from half a dozen languages acquired over a couple of millennia. The oldest of these loans date from the Greek occupation of Bactria in the third century BC, e.g. *mečán* (feminine) 'hand-mill, quern' taken over from *mēkhanḗ* at a time when *kh* was still an aspirated *k*, or *mačóγna*, *mačnóγza*, *mačlóγza* 'sling', which may be evidence for a weapon called *manganiká* (cf. Arabic *manǰanīq* 'mangonel') already at the same period. No special trace of a Zoroastrian or a Buddhist past remains, but the Islamic period has brought a great number of Arabic and Persian cultural words. Throughout the centuries everyday words also have been borrowed from Persian in the west and from Indo-Aryan neighbours in the east. Usually it is difficult to establish when: *maryalára* 'pearl', for example, could be from Greek *margarítēs*, or like it from an Old Persian **margāritā-*, or later from a Parthian or Sogdian form. Irregular assimilation makes it hard to decide when, say, *blárba* 'pregnant', *cerá* 'face, picture', *ǰalá* 'separate', *peǩ* 'happening' were acquired from Persian *bārbar*, *čihra*, *ǰudā*, *peš*, but it was long ago. The different stages of assimilation show that *žranda* 'water-mill' and *ǰandra* 'padlock' have been borrowed at different times from Lahnda (Western Panjabi) *ǰandar* 'mill' and *ǰandrā* 'padlock'. The sources of the many such Indian loanwords are particularly hard to distinguish. It is only when we come to *ǰarnáyl* 'general', *lâṙ* 'lord', *palṭə́n* 'platoon, regiment', *ṭikə́s* 'ticket, stamp' and *ṭwal* 'towel' that we are on firm ground again. The greater part of the basic vocabulary is nevertheless inherited Eastern Iranian. Still it is noteworthy how many original words have given way to neologisms. Most striking among these are some words for parts of the body: *γāx̌* 'tooth' (< **gaštra-* *'biter'), *stárga* 'eye' (< **stṛkā-* *'little star'), *təndáy* or *wəčwúlay* 'forehead' (the *tánda* 'thirsty' or *wəč* 'dry' part), *tóray* 'spleen' (the *tor* 'dark, black' organ), and several of unknown origin, such as *šā* 'back', *xwla* 'mouth'.

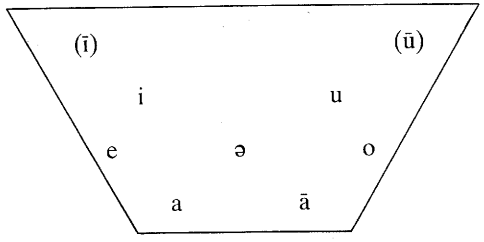
3 Phonology

The maximum inventory of segmental phonemes in Pashto is set out in table 26.1. Besides the common consonant stock of most modern Iranian languages, it comprises the dental affricates *c*, *j* [ts dz] and, thanks to its neighbourhood to Indo-Aryan languages, a set of retroflex, or cerebral, sounds. While the retroflex stops *ṭ*, *ḍ̂* occur only in loanwords, the *ṙ* has, as we have seen, also developed within Pashto. In distinction from the alveolar

trill *r* and from the dental (or alveolar) lateral *l*, it is basically a retroflexed lateral flap. Its nasal counterpart *n̂*, which does not occur word-initially, is a nasalised *r̂* — the nasalisation often extending to the preceding vowel — and not simply a retroflex nasal (which latter only occurs as an allophone of dental *n* before *t̂*, *d̂*).

Table 26.1: The Segmental Phonemes of Pashto

Vowels



Consonants

	Plosive	*Affricate*	*Fricative*	*Nasal*	*Lateral*	*Trill*	*Semi-vowel*
Bilabial	p b			m			w
Labio-dental			(f)				
Dental	t d	c j		n	l		
Alveolar			s z			r	
Retroflex	t̂ d̂		(x̌ ǧ)		n̂ r̂		
Post-alveolar		č ǰ	š ž				y
Velar	k g		x γ				
Uvular	(q)						
Glottal	(ʾ)		h				

The bracketed *f*, *q* and ʾ occur only in the elegant pronunciation of unassimilated loanwords from Persian and Arabic. Generally *f* is replaced by *p* (occasionally by *w*) and *q* by *k*, e.g. *fatīla* > *palitá* 'wick', *tafaḥḥuṣ* > *tapós* 'enquiry', *lafz* > *lawz* 'word, promise', *qiṣṣa* > *kisá* 'story', *qawm* > *kām* 'tribe'. The glottal stop (representing both Arabic *hamza* ʾ and *ʿayn* ʿ) is usually dropped, either without trace, e.g. *masʾala* > *masalá* 'question, matter', or having widened the adjacent vowel, as in *šarʿ* > *šára* 'holy law', *maʾmūr* > *māmúr* 'official', *šurūʿ* > *šuró* 'beginning', *mawẓiʿ* > *mawzé* 'place'. This resembles the treatment of word- and syllable-final *h*, *ḥ* in loanwords, e.g. *ṣaḥīḥ* > *sahí* 'correct', *fatḥ* > *fáta* 'victory', *iḥtirām* > *etərắm* 'respect', *makrūh* > *makró* 'abominable'.

Characteristic of Pashto are the two phonemes written *x̌*, *ǧ*. These developed originally as retroflex spirants [ŝ ẑ] and continue generally as such in the southwestern dialects, particularly the prestigious one of Qandahar, where they contrast with the post-alveolar *š*, *ž*. In the southeastern dialects this contrast has been lost. In most central dialects these phonemes are still realised distinctly, but as palatal spirants [x̌ γ̌]. In the north-east, however, they have coincided entirely with velar *x* and *g* (not *γ*!). The non-phonetic symbols *x̌*, *ǧ* thus represent a compromise between [ŝ/š/x̌/x] and [ẑ/ž/γ̌/g] respectively. This wide and striking variation between southwestern [paŝto] and north eastern [paxto] accounts for the description of the different dialects as 'soft' and 'hard' Pashto. It is noteworthy that the hard dialects, most directly exposed to Indo-Aryan influence, have also abandoned the dental affricates *c*, *j* (which lose their plosive element, to coalesce with *s*, *z*) and *ž* (which joins the affricate *ǰ*): in other words, with the exception of *x*, *γ* and *z*, their phonemic system has largely been Indo-Aryanised.

A notable feature of Pashto phonology, in which it differs from most other modern Iranian languages, is its toleration of groups of two or (including *w*) three consonants in word-initial position. Some hundred such groups occur, e.g. eleven with *š-* alone: *šp-*, *št-*, *šk-*, *šx-*, *šxw-*, *šm-*, *šn-*, *šl-*, *šr-*, *šr̂-*, *šw-*. Such initial groups are particularly unstable, being subject to various metatheses, assimilations and dissimilations. Thus *pẋa* 'foot', *kẋəl* 'pull' and *psarláy* 'spring' become hard *xpa*, *xkəl*, and *sparláy* respectively; *nwar* 'sun' occurs in different dialects as *nmar* and *lmar*, *rwaj* 'day' as *wraj*, *ǧmənj* 'comb' as *g(u)manj*, *mangáz*, and so on.

The vowel phonemes in table 26.1 are the stressed ones of standard Pashto, stress also being phonemic. The following diphthongs also occur: *ay*, *əy*, *āy*, *oy*, *uy*; *aw*, *āw*. The phonemic status of the historically long vowels *ī*, *ū* is questionable. In most dialects they have been reduced to coincide with *i*, *u*; i.e. length is here, as in the case of *e*, *o*, no longer significant but depends on position and stress. Stressed *a*, *ə*, are entirely distinct, e.g. *bal* 'alight': *bəl* 'other', *γla* 'female thief': *γlə* 'male thieves'. In unstressed position, however, they are usually in free variation. It is convenient to regard unstressed [a ə] both as allophones of *a*, i.e. to regard *ə* only as a strong- or weak-stressed phoneme. Otherwise (as is unfortunately the case in some modern works on Pashto, both Afghan and foreign) there are some dangers of confusion, for example in writing the diphthongs unstressed *ay* [~ əy] and stressed *ə́y*. In fact there is an important morphophonemic distinction between final *-áy*, *-́ay* and *-ə́y*. In the hard dialects *-ay* is generally monophthongised to an open [ɛ(:)], allowing *-əy* to shift and take its place at [ɛi]. In all dialects, but especially those of the south-west, there is a tendency towards regressive vowel harmony, in that the middle vowels *e*, *o* in syllables preceding high vowels *i*, *u* are themselves raised. Also in the south-west unstressed final *e*, *o* often coalesce with *i*, *u*, but not to the extent that morphological distinctions are lost. Thus *óse* 'you dwell' remains, in contrast to *ósi* 'he dwells'. *mor*, oblique *móre* 'mother', however, becomes *móri* [mu:ri], though still without rhyming with *lur*, obl. *lúre* 'daughter' > *lúri*. In some non-standard mountain dialects of the Afghan-Pakistan borderland, particularly of the Afridi and Wazir tribes,

there is a vowel shift of *ā* to [ɔ:], *o* to [œ: > ɛ:], and *ū* to [i:] (but not *u* > *i*); e.g. Waziri [plɔ:r] 'father', [mɛ:r] 'mother', [li:r] 'daughter'.

Three degrees of stress can be recognised: strong, medium and weak. Strong stress is comparatively free, in that it can occur on any syllable of a word, but it is mainly restricted to the first, last or penultimate syllables. It can also, particularly in verbal inflection, be mobile, though the shifts involved follow regular patterns, e.g. from *prewatə́l* 'to fall', also 'they (masculine) were falling', *préwatəl* 'they fell' and *prewátay* 'fallen (masculine singular)'. Occasionally lexical items may be distinguished solely by stress, e.g. *áspa* 'mare' : *aspá* 'spotted fever', *gorá* 'fair-skinned, European' : *góra* 'look!', *palitá* 'wick' : *palíta* 'indecent woman', *wāṙə́* 'small (masculine plural)' : *wā́ṙa* [-ə] 'all'.

4 Script

The earliest authenticated records of Pashto as a literary language date from the late sixteenth century, at a time when the whole area was, if turbulently, a part of the Mogul empire. The language has always been written in the

Table 26.2: Pashto Alphabet, with Transliteration

*	ا		ā medial	س	س		s
	آ		ā initial	ش	ش		š
	ب	ب	b	ښ	ښ		x̌
	پ	پ	p	[ص	ص		ṣ]
	ت	ت	t	[ض	ض		ẓ]
	ټ	ټ	ṫ (P also Urdu ٹ)	[	ط		ṭ, occasionally for ṫ]
[	ټ	ټ	ṣ]	[	ظ		ẓ]
	ج	ج	j	[	ع	ع	ʿ]
	چ	چ	č		غ	غ	γ
	څ	څ	ǰ (A خ)		ف	ف	f
	څ	څ	c	[	ق	ق	q]
[	ح	ح	ḥ]		ک	ک	k
	خ	خ	x		ګ	ګ	g
	د	د	d		ل	ل	l
	ډ	ډ	đ (P also Urdu ڈ)		م	م	m
[	ذ	ذ	ẓ		ن	ن	n
	ر	ر	r		ڼ	ڼ	n̂ (A ڼ)
	ړ	ړ	ṙ (P also Urdu ڑ)	*		و	w
	ز	ز	z	*	ه	ه	h
	ژ	ژ	ž	*	ی	ي	y
	ږ	ږ	ǧ				

Note: *On the function as vowel carrier of ا and ه in word-initial and final position respectively, and of و and ی medially and finally, see the discussion in the chapters on Arabic and Persian and table 26.3.

Perso-Arabic script (see the discussion of script in the chapters on Arabic and Persian), with the addition of certain modified letters to represent the peculiar consonant phonemes of Pashto. In the earliest manuscripts, from the late seventeenth to early eighteenth century, there is considerable variety in the representation of these consonants, but later a standard system emerged which persisted until recently. Since the adoption of Pashto as a national language in Afghanistan a number of innovations have been introduced into the script, which in the main make for more clarity. In Pakistan, on the other hand, there have been some tendencies, e.g. the occasional use of Urdu forms of letters and the phonetic representation of hard dialect forms (*ǧ* as *g*, *x̌* as *x*, *j* as *z* etc.), causing a departure from the classical standard. In table 26.2 the standard alphabet is given, with the modern Afghan (A) and Pakistani (P) forms as variants. The letters in square brackets occur only in unassimilated Arabic loanwords and the diacritics used in the transliteration are merely for mnemonic purposes, and have no phonetic significance. Thus ذ *ẓ*, ض *ẓ*, ظ *ẓ* are all pronounced [z], i.e. are all allographs of the phoneme *z*, usually written ز .

The Perso-Arabic script is by nature a consonantal one. The means by which the relatively simple vowel systems of Arabic and Persian are represented in it are inadequate for Pashto, where vowel representation is thus somewhat complicated: see table 26.3. The short vowels *a*, *ə* are not normally written, but are represented notionally by the superscript signs ◌́ *zwar* for *a*, ◌̄ *zwar-akay* for *ə*. In standard script the latter is sometimes represented by the sign ◌ٔ *hamza*, e.g. زهٔ *zə* 'I'. The signs ◌ِ *zer* and ◌ُ *peš* can represent *i* or *e* and *u* respectively, though all these vowels may also (particularly in Afghan practice) be written *plene* with the appropriate semi-vowel letters ی and و respectively; e.g. *injár* انځر or اینځر 'fig', *kisá* قصه or کیسه 'story', *de* دِ or دې 'your', *gul* ګل or ګول 'flower'.

Table 26.3: Vowel Representation

	Initially	*Medially*	*Finally*
a	اَ	◌َ	ه
ā	آ	ا	ا
ə	-	◌ّ	ه (P ۀ)
e	اِ	◌ِ	ې (P ے)
			(P ◌ِ in particles)
ay	اَي	◌ِ	ي (P ے)
əy	-	-	ئی (A ےnominal, ءیverbal)
i	اِ	◌ِ (A ◌ٖ)	-
ī	ایـ	◌ِ	ي
o	او	و	و
aw	او	و	و
u	اُ	◌ُ (A و)	و (P ◌ٗ)
ū	او	و	و

5 Morphology

Although it has departed considerably from the morphological patterns of Old and even Eastern Middle Iranian (as evidenced, for example, by Sogdian and Khotanese Saka) Pashto has nevertheless a remarkably complex nominal and verbal morphology. Two grammatical genders (masculine and feminine) and two numbers (singular and plural) are distinguished in both noun and, in part, verb. Although the nominal case system has essentially been reduced to a contrast between direct and oblique, there is in the singular also a vocative and a second oblique case used in conjunction with certain prepositions. Moreover the formatives used are not, as in practically all other still inflectional Iranian languages, restricted to suffixes. Alterations of stem vowels and stress and the substitution of endings also come into play.

Old Iranian masculine stems in *-a*, *-i*, (*-u*) have generally lost their final vowel, to appear in Pashto as consonant stems: *kāra-* > *kor* 'house, family', *gauša-* > *γwağ* 'ear', **gari-* > *γar* 'mountain'. The old feminine stems in *-ā* alone have survived practically unscathed as *-a* stems: *aspā-* > *áspa* 'mare', *uštrā-* > *úxa* 'she-camel', *wanā-* > *wána* 'tree', *xšapā-* > *špa* 'night'. Old *-an-* stems similarly preserved their nominative singular *-ā* to emerge as masculine nouns in *-a*: **maiθman-* > *melmá* 'guest'. Feminine stems in *-ī*, (*-ū*) also lost their final vowel, e.g. *hapaθnī-* > *bən* 'co-wife', **raθī-* > *lār* 'way, road', **witasti-* > *wlešt* 'span', but generally they adopt an *-a* from the general feminine form: **sraunī-* > *x̌n-a* 'buttock, leg', **strī-čī-* > *x̌áj-a* 'woman', **wahunī-* > **wēn* > *wín-a* 'blood', **zanu-* > *zə́n-a* 'chin'. Neuter stems joined either masculine or feminine, in the latter case also generally adopting a final *-a*: *raučah-* > *rwaj* f. 'day', **asru-* > *óx̌-a* 'tear', **gauna-* > *γún-a* 'colour', **parna-* > *pā́n̂-a* 'leaf'. Only rarely do old masculines become feminine, e.g. *angušta-* > *gút-a* 'finger', *safa-* > *sw-a* 'hoof'. Several forms in *-ya-*, nominal or adjectival (including the comparative in *-yah-*) yield Pashto *-ə*: **(p)tṛwya-* > *trə* 'paternal uncle', **t(a)igriya-* > *terá* 'sharp', *srayah-* 'better' > *x̌ə* 'good', **abrya-* > *orá* 'cloud'. A more common formative, however, as in Sogdian and Khotanese Saka, was the suffix *-ka-*. The resulting stems in *-aka-*, *-ika-*, *-uka-* became, via nominative or genitive **-ai* (as in Khotanese), either stressed or unstressed *-ay*. The feminine equivalent, originally **-akī-*, became *-áy* when stressed but *-e* when not: **daru-ka-ka-* > *largáy* 'wood', **sarda-ka-* > *sarấy* 'man', **spaka-* > *spay* 'dog' : **spakī-* > *spəy* 'bitch', **āsu-kī-* > *(h)osáy* 'deer', **náwa-ka-* > *náway* m. 'new' : **náwa-kī-* > *náwe* f. 'new'. The result of these far-reaching changes was three main masculine stem-types, ending in a consonant, stressed *-áy* or unstressed *-ay* respectively, and three corresponding feminine stem-types, ending in (generally unstressed) *-a*, stressed *-áy* or unstressed *-e*. There are also several exceptions which fit into this scheme as best they can, e.g. masculines ending in *-ə*, *-ā*, *-ū* and feminines in a consonant, *-ā*, *-e*, *-o*, all unchanged in the singular but approximating to the masculine consonant or feminine *-a* declension in the plural, or again masculines (professions) and feminines (abstracts) in *-i* joining the *-áy* and *-áy* stems respectively. The stem-types pair up in the case of adjectives to form the three declensions numbered 1, 4, 5 in the chart of adjectival declension. In all adjectival declensions the oblique singular forms are identical with the direct plural. Only nouns generally distinguish plural forms by plural markers, of bewildering variety. The 'prepositional' case is marked in the masculine by an unstressed *-a*, which probably represents an old ablative ending *-āt*, added to the direct case stem. In the feminine it coincides with the direct case. The vocative coincides in most, but not all, masculine singulars with the prepositional form and in most feminines with the oblique. The oblique, and also vocative and prepositional, plural marker *-o* (in soft dialects, stressed *-ó*, unstressed *-u*) is common to all declensions.

Adjectival Declension

	1 'other'	*2* 'ripe, cooked'	*3* 'bitter'	*4* 'thin, narrow'	*5* 'new'
Masculine					
Singular					
Direct	bəl	pox	trix	naráy	náway
Vocative	bála	póxa	tríxa	naráya	náwe
Prepositional	bála	póxa	tríxa	naráya	náwi
Oblique	bəl	pāxə́	tarxə́	narí	náwi
Plural					
Direct	bəl	pāxə́	tarxə́	narí	náwi
Oblique (Voc., Prepl.)	bálo[2]	paxó	tarxó	narío[2]/naró	náwyo[2]/náwo[2]
Feminine					
Singular					
Direct	bála	paxá	tarxá	narə́y	náwe
Vocative	bále[1]	paxé	tarxé	narə́y	náwe[1]
Prepositional	bála	paxá	tarxá	narə́y	náwe[1]
Oblique	bále[1]	paxé	tarxé	narə́y	náwe[1]
Plural					
Direct	bále[1]	paxé	tarxé	narə́y	náwe[1]
Oblique (Voc., Prepl.)	bálo[2]	paxó	tarxó	narə́yo[2]/naró	náwyo[2]/náwo[2]

Note: Qandahari: [1]báli, náwi. [2]bálu, naríu, náw(y)u.

There are also two further types of consonant stem (declensions 2, 3), represented among both nouns and adjectives, in which stress and vowel changes occur which may go back to a very early stage of the language. In the first type, comprising some (but not all) monosyllabic nouns and adjectives

with the stem vowel *o* or *u* and some nouns with final *-un*, the oblique singular and direct plural masculine substitute the vowel *-ā-*, and the oblique plural and entire feminine the vowel *-a-*, all with additional stressed endings. In the other type the same stressed endings occur with a stem either unchanged or with the stem vowel reduced to an *-a-* or nil. Thus *kuṇ* 'deaf' has the plural *kāṇə́* and feminine *kaṇá*, but *ruṇ* 'light' plural *ruṇə́*, feminine *ruṇá*; *soṛ* 'cold', plural *sāṛə́*, but *sur* 'red' plural *srə*. Similarly declined are a few words ending in stressed *-ə*: *x̌ə* 'good', singular and plural masculine, *x̌a* feminine singular, *x̌e* plural. A last set of adjectives comprises all those which end in any other vowel — *a*, *ā*, *e*, *i*, *o*, *u*. These are indeclinable for number, gender or case, except that they may take the universal oblique plural *-o*.

The plural of masculine nouns of the first declension, which also includes those ending in *-ə*, *-a*, *-u*, is generally *-úna*, oblique *-úno*, e.g. *lās* 'hand', *lāsúna*, *zṛə* 'heart', *zṛúna*. Animate nouns take the suffix *-ắn*, borrowed from Persian, oblique *-ắno*, e.g. *ux̌* 'camel', *ux̌ắn*, *lewə́* 'wolf', *lewắn*; before this suffix a *-y-* is inserted after *-ā*, e.g. *mullāyắn* 'mullahs', or a *-g-* after other vowels, e.g. *nikəgắn* 'ancestors'. Inanimate nouns in *-u* take the same ending: *bāṇugắn* 'eye-lashes'. Feminine nouns of this declension ending in a consonant or *-a* behave like adjectives even in the plural, e.g. *lār* 'road', plural *lắre*, *xwla* 'mouth', *xwle*. Animate ones ending in *-o*, however, take the mixed Persian and Pashto suffix *-gắne*, e.g. *pišogắne* 'cats', and those in *-e* change this to *-yắne*, e.g. *xwắx̌e* 'mother-in-law', *xwāx̌yắne*. Inanimate feminine nouns in *-ā*, *-o* on the other hand take an unstressed plural ending *-we*, e.g. *mlắwe* 'waists'. Nouns of declension 2 generally follow the adjectival pattern, e.g. *sor* 'rider', direct plural *swārə́*, oblique *swaró*, *pax̌tún* 'Pashtun', plural *pax̌tānə́*, feminine *pax̌taná* 'Pashtun woman', etc. Some such nouns, however, follow declension 1 in the plural, e.g. *žwandún* 'life, livelihood', oblique singular *žwandānə́*, plural *zwandunúna*. This is also the case with declension 3: *γar* 'mountain', plural *γrə* or *γrúna*, *trə* 'paternal uncle', *trə* or *trúna*. A number of nouns which only modify the vowel of their final syllable can also be classed here: *melmə́* 'guest', plural *melmə́* (or *melmānə́*), *duxḿán* 'enemy', *duxḿən*. A few nouns ending in *-ba* (sometimes alternating with *-bun*) follow declension 3 in the singular and 2 in the plural, e.g. *γobə́* (or *γobún*) 'cowherd', oblique singular *γobə́* (*γobānə́*), plural *γobānə́*, *γobanó*. Nouns of declensions 4 and 5 also follow the adjectival pattern, except that animates may also take the appropriate *-ān* ending, e.g. *spay* 'dog', plural *spi* or *spiắn*, *spəy* 'bitch', *spəy* or *spiắne*, *buḍ̂əy* 'old woman', *buḍ̂əygắne* or *buḍ̂yắne*. Even this catalogue does not exhaust the full variety of plural forms. The class of nouns of relationship is particularly rich in irregularities, as the following list will show: *plār* 'father', plural *plắrúna*; *mor* 'mother', *máynde* (*mándi*); *xor* 'sister', *xwáynde* (*xwándi*); *tror* 'aunt', *tráynde* (*trándi*), *troryắne*; *yor* 'husband's brother's wife', *yúṇe*; *lur* 'daughter', *lúṇe*; *wror* 'brother', *wrúṇa*; *wrārə* 'brother's son', *wrerúna*; *zoy* (*zuy*) 'son', *zāmən*.

Several nouns, particularly those denoting substances, occur only in the plural, whether masculine, e.g. *čars* 'hashish', *γanə́m* 'wheat', *γwaṛí* 'cooking oil', *māγzə́* 'brain', *oṛə́* 'flour', *tambākú* 'tobacco', *wāx̌ə́* 'grass', or feminine, e.g. *čāy* 'tea', *obə́* 'water', *orbə́še* 'barley', *šomlé* 'buttermilk'. To these may be added words with a collective meaning, such as *xalk* 'people', onomatopoeics ending in *-ahắr* denoting noises, e.g. *šrapahắr* 'splashing' and all verbal infinitives used as nouns. A last quirk of nominal declension concerns masculine consonant stems, mostly inanimate, when qualified by and directly following a cardinal number higher than 'one', or a similar adjective such as *co* 'several, how many?'. Instead of appearing in the plural, as all other nouns then do, they take a 'numerative' ending *-a* in the direct case. This also affects the higher numbers (*šəl* 'score', *səl* 'hundred', which then takes the form *saw*, *zər* 'thousand') and the enumerative words which frequently appear between number and noun: *co jə́la* 'how many times?', *dre kắla* 'three years', *calór sáwa saṛí* 'four hundred men', *pinjə́ zára míla* 'five thousand miles', *atə́ kitắba* or *atə́ ṭuka kitābúna* 'eight (volumes) books'. This numerative ending may well be a last relic of the ancient dual.

The direct case of nouns serves both for the grammatical subject and direct object of verbs. Case relationships are all expressed by pre- and postpositions or a combination of both, used with one of the oblique cases: an oblique form alone may have adverbial sense, e.g. *yáwa wráje* 'one day'. The simple prepositions are *da* 'of', which provides the only means of expressing a genitive or possessive relationship, *la* 'from', *pa* 'in, at etc.', *tar* 'to, from': postpositions, appearing independently or in combination with prepositions, are *na* 'from', *ta* 'to', *bắnde* 'on', *cáxa* and *jáne* 'from', *kx̌e* (generally reduced to *ke*, *ki*) 'in', *lắnde* 'under', *larə́* 'for', *pās* 'above', *pasé* 'after', *póre* (*púri*) 'up to', *sarə́* 'with'. Combinations of pre- and postpositions vary somewhat from dialect to dialect: common examples are *da... na* 'from', *la... sarə́* 'with', *pa... kx̌e* 'in', *pa... bắnde* 'on', *tar... póre* 'up to, till'. Most pre- and all postpositions take the main oblique case. The second oblique case, which as it serves no other function can for convenience be called the 'prepositional' case, is as a rule taken only by the simple prepositions *be* 'without', *la* and *tar* and by *pa* (*...kx̌e*), but this last, remarkably, with feminine nouns only.

With pronouns things are somewhat different. Pashto has, in fact, comparatively few independent pronouns. Besides those for the first and second persons, singular and plural, there are proximate and remote demonstrative pronouns, which double for the third persons, and a few indefinite and interrogative forms. For the rest paraphrase is used, much as in English. e.g. *jan* 'body, self' for 'my-, your-, himself etc.', *yaw... bəl* 'one... other' for 'each other'. The place of a relative pronoun is taken by the conjunctive particle *če* 'that', '(the man) who came' being expressed as 'that he came', and 'whose house...' as 'that his house...' and so on.

Pronouns

	Singular		*Plural*			*'who?, somebody'*	*'what?, something'*
	1	2	1	2			
Direct	zə	tə	muǧ[1]	tấso	(tấsi)	cok	cə
Oblique	mā	tā	muǧ	tấso	(tấsi)	čā	cə
Possessive	jmā	stā	jmuǧ[1]	stấso	(stấsi)	da čā	

	'this'		*'that'*
Masculine			
Direct	day	dáγa	háγa
Oblique	də	dáγə	háγə
Feminine			
Direct	dā	dáγa	háγa
Oblique	de	dáγe	háγe
Plural			
Direct	duy	dáγa	háγa
(Personal)		dáγuy	háγuy
Oblique	duy, dío	dáγo	háγo

Note: [1] Hard dialects, mung, zmung.

Of those pronouns which show a difference, the first and second person singular ones are unique in that the direct forms act only as subject, the oblique case forms (distinct only in the singular) being used both for the direct and a prepositional object. The personal pronouns also have distinct possessive forms, combining the old preposition *hača* 'from' in the form *j-*, (*z-*), *s-*, which may also occur with postpositions usually combined with *da*, e.g. *jmā na* 'from me'. There are also two kinds of pronominal particle, one independent and one enclitic. The enclitics are only incompletely distinguished for person and number: 1st singular *me*, 2nd singular *de*, 3rd singular and plural *(y)e*, 1st and 2nd plural *mo*. They fulfil all the oblique functions of the pronouns except that of prepositional object, though even in this case there are traces of the third person form to be seen in combinations of the sort of English 'therefrom, -on, -in', Pashto *tre* < **tar-e*, *pre* < **par-e*, *pakže* < **pa kžé-ye*. The independent forms, *rā*, *dar*, *war*, are by origin local adverbs 'hither, thither' and 'yonder' and still act as such when no person is involved. They come to act as pseudo-pronouns, however, distinguishing only person, neither number nor gender. Thus they may be governed by post- but not prepositions, e.g. *dar sara* 'with you', or serve as a prepositional object with certain verbs: *war ba nənawázəm* 'I shall enter therein' or 'go in to him', according to context.

The verbal morphology of Pashto, as with all other modern Iranian languages, is based on the opposition between two stems, one present and one past. Present stems are either simple (inherited or borrowed ones) or secondary (made with the formatives *-eǧ-* intransitive or *-aw-* transitive and causative). These latter both generally form denominatives (*num-eǧ-* 'be named') or serve to assimilate loan-words (*bah-eǧ-* 'flow', from Hindi *bah-nā*), but in some cases *-eǧ-* also distinguishes a continuous sense from a timeless or habitual one: *dəlta dera wāwra óri* 'here much snow falls (lit. rains)' : *oréǧi* 'it is raining'. The past stems are essentially old perfect passive participles in *-ta-*, though more often than in any other Iranian language phonetic developments have disguised the characteristic dental ending. In contrast, for example, to Persian *sūz-ad*, *sūxt* 'it burns, burnt', Pashto has *swaj-i*, *su*. A dental may even arise in the present and disappear from the past, e.g. *təxt-* 'flee' < **tṛsa-*, against *təx* 'fled' < **tṛšta-*, or the two stems may coincide, as in *ačaw-* 'throw' < **ā-škaba-* and *-škafta-*. As a result a new past marker has emerged, a stressed *-ál-*, identical with the infinitive ending *-ál* (<**-ati-*), which is added to the past stem whenever the need is felt to arise. Corresponding to the intransitive present formative *-eǧ-*, and generally but not always paired with it, there is a past formative *-ed-*.

On the basis of these two stems simple tenses are formed by the addition of personal endings, stressed or not according to the stem, which distinguish first and second persons singular and plural, but third person only, without difference of number. Thus, from *lwedál* 'fall' and *ačawál* 'throw' are formed the present and past paradigms shown here.

		Present		*Past*	
Singular	1	lwéǧ-əm	acaw-ám	lwéd-əm	ačaw-ál-əm
	2	lwéǧ-e	acaw-é	lwéd-e	ačaw-ál-e
	3 m.	lwéǧ-i	acaw-í	lwéd(-á)	ačāwá
	3 f.			lwed-ála	ačaw-ála
Plural	1	lwéǧ-u	acaw-ú	lwéd-u	ačaw-ál-u
	2	lwéǧ-əy[1]	acaw-áy[1]	lwéd-əy[1]	ačaw-ál-əy[1]
	3 m.	lwéǧ-i	acaw-í	lwed-ál	ačaw-ál
	3 f.			lwed-ále	ačaw-ále

Note: [1] Qandahari, 2nd plural *-āst*, thus *lwéǧ-āst* etc.

The original composition of the past tense, from a passive participle and the copula, is still clear in the third person, where the copula is lacking and the forms are declined like adjectives, though frequently with an irregular masculine singular form in which a stem vowel *-a-* is lengthened to *-ā-* or changed to *-o-* (*xatál* 'rise', *xot* 'rose'). Moreover the old participle of transitive verbs, as past stem, retains its passive meaning throughout: *ačawám* 'I throw', but *ačawálam* 'I was being thrown'. This is also true of the modern past participle, a regular adjective of declension 5, e.g. *lwedálay* 'fallen', *ačawálay* '(having been) thrown', which with the auxiliary verb 'be' forms periphrastic tenses. The modern copula similarly betrays the probable pronominal origin of its third person forms. The simple perfect, for example, is formed as in the chart given here.

		Masculine	*Feminine*	*M./F.*
Singular	1	lwedálay yəm	lwedále yəm	ačawəlay/e yəm
	2	lwedálay ye	lwedále ye	ačawəlay/e ye
	3	lwedálay day	lwedále da	ačawəlay/e day/da
Plural	1	lwedáli yu	lwedále yu	ačawəli/e yu
	2	lwedáli yəy	lwedále yəy	ačawəli/e yəy
	3	lwedáli di	lwedále di	ačawəli/e di

'I have fallen' etc., but 'I have been thrown', etc. In contrast to the present tenses, 'I throw it' etc., there is thus no means of expressing the active non-present tenses of the transitive verbs by forms in concord with a logical subject or agent in the direct case. Instead of 'I threw it', therefore, an ergative construction is obligatory, which — to avoid the passive 'it was thrown by me' — can only be expressed in English as 'me thrown it'. In Pashto the logical object but grammatical subject, inherent in the verb, may of course be expressed by an independent form, but if it is pronominal it need not be. The agent, however, must appear, in the oblique case. A personal pronoun may then be represented either by an independent form (*mā* etc.), which then generally precedes the grammatical subject, or by an enclitic (*me*, etc.). Various different possible paradigms thus arise (a matter to which we shall return), e.g.:

mā kāñay...	or *kāñay me ačawálay day*	'I have thrown a stone',
tā zə...	or *zə de ačawálay yəm*	'you have thrown me',
hayə ačawálay day	or *ačawálay ye day*	'he has thrown it'.

In contrast to this a real passive usually only occurs when the agent is unknown or at least not expressed. Such a passive is formed by the past participle, or in soft dialects the 'old past participle', i.e. the third person past forms, with the auxiliary verb *kedəl/šwəl* 'become': *ačāwə́/ačawálay keğəm* 'I am being thrown', *ačawála/ačawále šwa* 'she was thrown', *ačawə́l/ačawəli šə́wi di* 'they have been thrown'. A full passive, with the agent expressed by a prepositional phrase like 'by means of', as in *kāle če da nāwe la xwā roγ šəwe wi* 'clothes which will have been made by (lit. from the side of) the bride', is a rarity.

Pashto employs two further means, besides the different temporal stems, for distinguishing a series of forms which intricately mark differences of mood and aspect. The one means is to provide each verb with secondary stems, present and past II. This is mostly done by means of a stressed separable prefix *wə́* (eastern *(w)u*), e.g. *wə́lweğ-*, *wə́lwed-*. With an initial *a-* the prefix forms *wā-*, which then makes itself independent of the verb as a pseudo-preverb, e.g. *wā́čaw-*, *wā́čawəl-*. True preverbs, like *kxe* and *nána* 'in', *póre* 'to, across', *pre* 'off, from', exclude the prefix *wə́*. Instead they attract the stress to themselves, e.g. from *kxewatə́l* 'enter', present stem I *kxewə́z-*, II *kxéwəz-*, past II *kxéwat-*. Half a dozen of the commonest verbs

combine stems of widely different origins, so that the I and II stems are sufficiently distinct to dispense with the help of *wə*. Among these are *kedə́l* 'become', present I *kéğ-*, II *š-*, past II *šw-*; *kawə́l* 'do, make', present I *kaw-´*, II *k(r̂)-´*, past II *kr̂-*; and the particularly complicated *tlə́l* 'go', present I *j-*, II *wlār̂ š-*, past II *wlār̂-*, but *rā-tlə́l* 'come (hither)', present I *rā-j-´*, II *rā́-š-*, past II *rā́-γl-*, which follows the same pattern with alternative prefixes in *dar-tlə́l* 'come, go to you', *war-tlə́l* 'go to him'. Denominative verbs distinguish their I and II stems in yet another way. Here the composite primary stems are opposed to secondary stems in which the independent inflected nominal form is compounded with the secondary stems of *kedəl* or *kawəl*: thus from *ǰor̂* 'well, ready, agreeable', *ǰor̂edə́l* 'get well, be made, made ready, agree', present I *ǰor̂éğ-*, II *ǰór̂ š-*, past II *ǰór̂ šw-*. The contrast is even more marked with words of declension 2 or 3, since they form denominatives from the 'weak' feminine stem, e.g. from *pox* 'cooked, ripe', *paxawə́l* 'cook', present I *paxaw-´*, II *póx k(r̂)-*, past II *póx kr̂-*.

The other means is a movable enclitic particle *ba*. Its movements call to be described below, but for the moment we shall consider it in relation to the finite verb alone. It remains only to mention the distinctive endings of the imperative (singular *-a*, plural *-əy*) and of the conditional mood (*-āy*, eastern *-ay*, for all persons) and we have all the ingredients for the first part of the verbal system sketched in table 26.4. The lower part comprises both the periphrastic tenses, formed from the past participle, and the forms expressing the potential mood, which are compounded of the simple conditional form and the auxiliary verb *šwəl* (Qandahari *swəl*) 'be able', the forms of which chance to be identical with the secondary ones of *kedəl* 'become'. Here the prefix *wə* seems to have lost its significance, to become facultative.

Between the present I and II there is a difference of mood, I being indicative, 'falls, is falling', II subjunctive, '(that, if) it fall'. In the corresponding future forms, however, with the addition of the particle *ba*, there is a distinction of aspect, I being durative, 'will be falling', II perfective, 'will fall'. This holds good also in part for the imperative, I 'keep on falling', II 'fall'. But the prohibitive, with the particle *ma* 'not', cuts across this. It is normally only formed from stem I, regardless of aspect: *má lweğa* 'do not fall'. The past II is again perfective, 'fell', in contrast to the past I with durative sense, 'was falling', or occasionally inchoative, 'was about to fall'. The addition of *ba* in this case, although giving a sense of customariness, does not entirely remove the aspectual distinction: III 'used to fall, be falling, continuously' : IV 'used to fall repeatedly'. With the conditional forms I and II no aspectual difference can be seen: both can express present or future conditions, '(if) it were falling' or 'were to fall', the possible consequences '(then) it would fall' being expressed either by the past III or IV, or the conditional III (IV being unusual). The periphrastic tenses are by nature all perfective. With the perfect forms the sense follows that of the

Table 26.4: The Verbal System

Present I lwéği	Present II wəlweği	Future I lwéği ba	Future II wə-ba-lweği
Imperative I lwéğa	Imperative II wəlweğa		
Past I lwedə́	Past II wəlwed	Past III lwedə́ ba	Past IV wə-ba-lwed
Conditional I lwedǎy	Conditional II wəlwedǎy	Conditional III lwedǎy ba	
Perfect I lwedə́lay day	Perfect II lwedə́lay wi		Future Perfect lwedə́lay ba wi
Past Perfect I lwedə́lay wə		Past Perfect III lwedə́lay ba wə	
Perfect Conditional I lwedə́lay wǎy		Perfect Conditional III lwedə́lay ba wǎy	
Potential Present (wə́)lwedǎy ši		Future (wə́)lwedǎy ba ši	
Past (wə́)lwedǎy šu		Past III (wə́)lwedǎy ba šu	
Conditional (wə́)lwedǎy šwǎy			

auxiliary verb, i.e. between perfect I and II there is a difference of indicative, 'has fallen', and subjunctive, '(if) it (should) have fallen', in the third person only, as the other persons of the copula have common forms for both I and II. The future perfect only occurs in the II form, there being no durative future form of the copula. It has both senses of the corresponding English tense, 'it will (i.e. must) have fallen (by now, or some past time'), or 'it will have fallen (by some future time)'. The perfect conditional I expresses no longer possible conditions, '(if) it had fallen', and the past perfect III or the perfect conditional III the consequence, '(then) it would have fallen'.

6 Syntax

The first important syntactic feature to be considered is word order, which, starting from the noun phrase, is fairly inflexible in Pashto. All qualifiers precede the head of a noun phrase. The English freedom to say 'that man's

hand' or 'the hand of that man' is denied a Pashto-speaker, who has only *da hayə saṙi lās* 'of that man hand'. Missing is an article in Pashto, though this lack may occasionally be made up by the use of a demonstrative or the word *yaw* 'one'. Combining *yaw zoṙ kə́lay* 'an old village' and *tange kucé* 'narrow streets' yields *da yawə zāṙə kə́li tange kucé* 'an old village's narrow streets'. Only the personal possessive forms can precede the *da* group: *stə́so da kə́lo kucé* 'your villages' streets'. The apparent parallelism breaks down, however, when the noun phrase is governed by a pre- or postposition. The postposition appears at the end of the entire phrase, but a lone or accompanying preposition must be placed immediately before the head and its attributes. Thus 'from the very narrow streets of your old villages' can only be *stə́so da zaṙo kə́lo la ḋero tango kucó na* 'your of-old-villages from very-narrow-streets-from'.

Since both subject and direct object of a non-past transitive verb appear in the direct case, only a fixed word order can disambiguate them. Pashto has therefore become an inflexible subject–object–verb language: *saṙáy x̌əja wíni* 'man woman sees' can only mean 'the (a) man sees the (a) woman'. The positioning of adverbial phrases is freer. The order of the following sentence seems to be the most natural one: *(A:hara wraj) (B:pa kum waxt če kəli ta ji) yaw saṙay (C: pa ḋer tāǰub) yawa barbanḋa x̌əja (D: pa lāra kx̌e) wini* '(every day) (at what time he goes to the village) a certain man (to his great surprise) sees a naked woman (on the road)'. But an alternative arrangement *(A) (C) yaw saṙay (B) (D) yawa barbanḋa x̌əja wini* is just as thinkable as the English '(A), (C), a certain man, (B), sees (D) a naked woman'. Given the inflexibility of the SOV order in the non-past, it is not surprising that the ergative construction of the past parallels it. With independent forms the necessary word order is agent–patient–verb or, translated into terms of grammatical concord, agent (oblique)–subject (direct)–verb (concord): *mā saṙáy wə́lid* 'I saw the man', *saṙí x̌əja wə́lidəla* 'the man saw the woman', *zaṙo kə́lo ba tange kucé larə́le* 'old villages used to have narrow streets'. This simple rule is disturbed, however, by the fact already noted that a pronominal agent may be expressed by an enclitic form, and enclitics are a law unto themselves in Pashto.

Besides those already met, pronominal *me*, *de*, *(y)e*, *mo* and verbal *ba*, Pashto has a few more enclitics. *de (di)* may lose its original pronominal force and, as an ethic dative, simply give the present II (subjunctive) form a jussive sense: *kitābúna de rā̆wṙi* 'let him bring the books'. Then there are the conjunction *xo* 'but' and the adverb *no* 'so, then, still, yet', which can be used enclitically. Two or three of these may occur together, when they have the following fixed pecking order:

xo / ba / me, de, ye, mo / no

pré-xo-ba-ye-nə́-ğdəm 'but I shall not leave it', *dā̆-xo-ba-me nə́ kāwə* 'but

this I used not to do'. As a group they always seek the earliest possible support in a clause, namely the first syntagm, be it word, phrase or more, bearing at least one main stress. In short, when the agent is expressed by an enclitic pronoun its position is not relative to the grammatical subject at all, but is governed by the word order of the clause as a whole: *šikāyát-ye wákəř* 'complaint him made', i.e. 'he complained', *(da xéţe la xwáğ cəxa)-ye šikāyát wákəř* '(of stomach from pain-from) him complaint made', i.e. 'he complained of stomach ache', *hálta-ye (da... cəxa) šikāyát wákəř* 'there he complained (of stomach-ache)'. Conversely as the content of a sentence is reduced an enclitic agent is forced back until it may be supported by parts of the verb, including a preverb, alone: *paroskál-ba-mo xar rấwost/xár-ba-mo rấwost/rấ-ba-mo-wost* '(last year) we used to bring (the donkey) it'. All this is equally true of the enclitic pronouns in their other functions, as direct object or possessive: *nə́-ye važni* 'he does not kill it', *magar važnáy-ba-ye nə́* 'but kill them you shall not'; *(stā da xeţe ilấj kawa* or) *da xeţe ilấj-de kawa* 'have your stomach treated', *xayrāt pradáy wə, no xéţa-xo-de xpála wa* 'the free food was provided by somebody else, but the stomach was your own'. Even poetic licence and transpositions *metri causa* cannot affect the rule. Instead of prosaic **mine-ba-me laryun da tan kor səway wə, ka-me žařā pa himāyat nə rātlay* 'love would long since have burnt the house of my body, if weeping had not come to my support (in dousing it)', the poet 'Abdul Hamid Mohmand has:

da tan kór-ba-me laryún wə mine sə́way
ka-me nə́ rātlay žařā pa himāyát.

The only constituent that can hold an enclitic back from its natural support is a relative clause immediately following it. A clause is clearly felt to be too diffuse to support enclitics, which are forced to attach themselves to the next best, i.e. following, syntagm: *haya nǰəláy-me māxấm sinemấ ta byāyi* 'that girl is taking me to the cinema this evening', *haya nǰəláy, če os-mo wálidəla, māxấm-me sinemấ ta byāyi* 'that girl we just saw is taking me to the cinema this evening'. Sometimes, however, an enclitic may burst the bounds of its own subordinate clause to move to the front of the main clause, e.g. instead of *har sabā če yrə-ta-ba tə*, 'every morning, when he would go to the mountain', we find *har sabā-ba če yrə-ta tə*; instead of *pa har jāy-kxe če mumí-ye*, 'in whatever place he finds it' — *pa har jāy-kxe-ye če mumi*.

Of agreement in Pashto there is little to be said except that, where the forms permit it, it is all-pervading. Adjectives, whether attributive or predicative, agree in number, gender and case with their head nouns or subjects respectively: *zmā grấna aw mehrabấna plấra* 'my dear and kind father!' (masculine singular vocative), *kláka zmə́ka* 'firm earth', *zmə́ka kláka da* 'the earth is hard' (feminine singular direct), *če stā mlā sáma ši yā da nóro xálko mlāgấne kubáy ši* (they asked a hunchback whether he wanted)

'that your back should become straight (feminine singular direct) or other people's (masculine plural oblique) backs should become hunched (feminine plural direct)'. This agreement extends to adjectives used adverbially, e.g. *đer* 'much, many' but also 'very', *hawấ đéra tawdấ wi* 'the climate is (always) very hot' (feminine singular direct), *kixtáy-e kláka wə́niwəla* 'he siezed hold of the boat firmly' (feminine singular direct). While the agreement of subject and verb is normally restricted to person and number (note *Tor zə aw tə botlu* 'Tor took (1st plural) me and you'), with the third person singular copula gender also comes into play: *ấs day* 'it is a horse', *áspa da* 'it is a mare'. In the ergative construction, with all third person forms both gender and number are marked throughout: *xə́je ās wə́wāhə* 'the woman struck the horse', *áspa-ye wə́wahəla* 'he/she/they struck the mare', *āsúna-ye wə́wahəl* '...struck the horses', *áspe-ye wə́wahəle* '...struck the mares'. In the perfective forms of denominative verbs, in which the nominal element is free, agreement is naturally to be expected: *zə bāyad ywáxe paxé kəm* 'I must cook some meat (feminine plural direct of *pox*)'. More unexpectedly, even nouns forming denominatives become adjectivised in this context: thus from the Persian loanword *yād* 'memory', forming *yādedə́l* 'be remembered', we find *haya xə́ja-me yấda šwa* 'I remembered that woman'.

If we compare the archaic structure of Pashto with the much simplified morphology of Persian, the leading modern Iranian language, we see that it stands to its 'second cousin' and neighbour in something like the same relationship as Icelandic does to English.

Bibliography

The best modern study in English is Penzl (1955), despite minor errors; it is based on the work of Afghan grammarians. Trumpp (1873) remains, despite its age, the best grammar based on classical Pashto literature. For syntax, Lorimer (1915) is an amateur study, but a mine of information. Morgenstierne (1942) is a unique historical study, by the leading specialist.

References

Grjunberg, A.L. 1987. *Očerk grammatiki afganskogo jazyka (Pašto)* (Nauka, Leningrad)

Lorimer, D.L.R. 1915. *(A) Syntax of Colloquial Pashtu* (Oxford University Press, Oxford)

Morgenstierne, G. 1942. 'Archaisms and Innovations in Pashto Morphology', *Norsk Tidskrift for Sprogvidenskap*, vol. XII, p. 87–114

Penzl, H. 1955. *A Grammar of Pashto* (American Council of Learned Societies, Washington, DC)

Trumpp, Ernest. 1873. *Grammar of the Paštō or Language of the Afghans* (Trübner, London)