The World's Major Languages

Second Edition

Edited by
Bernard Comrie

The World's Major Languages

Second Edition

Edited by
Bernard Comrie

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First edition published in 1987 by Croom Helm
Reprinted with revisions and additions in 1989 by Routledge

Second edition published in 2009 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN
www.routledge.co.uk

Simultaneously published in the USA and Canada
by Routledge
270 Madison Avenue, New York NY 10016
www.routledge.com

Routledge is an imprint of the Taylor and Francis Group, an informa business

This edition published in the Taylor & Francis e-Library, 2008.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloguing-in-Publication Data
A catalog record for this book has been requested

ISBN 0-203-30152-8 Master e-book ISBN

ISBN 978-0-415-35339-7 (hbk)
ISBN 978-0-203-30152-4 (ebk)

Contents

30

Turkish and the Turkic languages

Jaklin Kornfilt

1 General and Historical Background

A strict terminological distinction should be drawn between Turkic, the name of a language family, and Turkish, the name of a language. Although Turkish is by far the largest language (in terms of number of speakers) in the Turkic family, it accounts for only some 30 per cent of the total number of speakers of Turkic languages. The main geographic locations of Turkic languages are: (1) Turkey (Turkish), (2) the former USSR and Iran: the Caucasus and northwestern Iran (e.g. Azerbaijani), formerly Soviet Central Asia, Kazakhstan and southern Siberia (e.g. Uzbek, Kazakh, Turkmenian, Kirghiz) and on the Volga (e.g. Tatar). One Turkic language (Yakut, or Sakha, as it is called by its native speakers) is spoken in northern Siberia. (More than one inhabitant in ten of the formerly Soviet areas is a native speaker of a Turkic language.) In addition, there are substantial Turkic-speaking communities in northwestern China (especially Uighur, and also Kazakh).

In terms of linguistic structure, the Turkic languages are very close to one another, and most of the salient features of Turkish described below (e.g. vowel harmony, agglutinative morphology, verb-final word order, nominalised subordinate clauses) are true of nearly all Turkic languages, with only minor modifications. This similarity of structure makes it difficult to determine the precise number of Turkic languages and their boundaries and to sub-classify them, since one typically finds chains of dialects, with adjacent dialects in essence mutually intelligible and mutual intelligibility decreasing as a function of distance, rather than clear language boundaries. Only one Turkic language, Chuvash, spoken on the middle Volga, is radically different from all its relatives.

The external genetic relationships of the Turkic family remain controversial. The most widely accepted affiliation is with the Mongolian languages (in Mongolia, northern China and parts of the former USSR) and the Tungusic languages (Siberia and northeastern China), to form the Altaic phylum; the typological similarities among these three families, though striking (e.g. vowel harmony, SOV word order typology) are not proof of genetic relationship, while even the shared vocabulary has been argued

to be the result of intensive contact rather than common ancestry. Bolder hypotheses would extend the Altaic phylum eastwards to include Korean, perhaps even Japanese; or northwards to include the Uralic family (to give a Ural-Altaic phylum).

Turkish is the official and dominant language of Turkey (Turkish Republic), where it is the native language of over 90 per cent of the population, i.e. some 70 million people. (The largest linguistic minority in the Turkish Republic is formed by Kurdish speakers, mainly in southeastern Turkey; small minority language communities are formed by speakers of Arabic, of some Caucasian languages, and, especially in the European part of the country, by speakers of Gagauz, a closely related Turkic language primarily spoken in Moldova.) Turkish is also a co-official language (with Greek) in Cyprus, where it is spoken by 18–19 per cent of the population, or about 140,000 people. But the largest number of Turkish speakers outside Turkey, perhaps close to one million, is to be found in the Balkans, especially in Bulgaria, but also in the former Yugoslavia (particularly in Macedonia) and in Greece.

Although there is no general agreement in Turkological literature on the most adequate geographic grouping of the Turkic languages, we shall go along with those sources that classify the contemporary language spoken in the Turkish Republic within a South-West (or Oγuz) group, together with Gagauz, Azerbaijani and Turkmenian, the latter forming the eastern component of the group. Within this group, some sources differentiate a subgroup called *Osman* (i.e. Ottoman), which would consist of the following dialects: Rumelian, Anatolian and South Crimean. Modern standard Turkish represents a standardisation of the Istanbul dialect of Anatolian.

The question of the ancestor language of this group is not settled, either. It seems established, however, that the language of the oldest documents (i.e. the Orkhun inscriptions and the Old Uighur manuscripts) is the ancestor of another group, namely of the Central Asiatic Turkic languages; the South-West languages are presumably descendants of the language of the 'Western Türküt' mentioned in the Chinese Annals.

The ancient languages of this group would be Old Anatolian and Old Osman. These labels themselves are misleading, however, and have more political and historical justification than linguistic motivation, since there are no clear-cut criteria to distinguish the languages they represent from one another – while there might be more reason to distinguish Old Osman (which is usually claimed to extend until the fifteenth century, ending with the conquest of Constantinople) from Ottoman proper; but, even there, no justification exists for a strict cut-off point.

The first Anatolian Turkish documents date from the thirteenth century and show that the literary tradition of Central Asia was only very tenuously carried over by the Turkish people (who had been converted to Islam earlier) after invading Anatolia from the east in the late eleventh century. It is clear that these tribes were influenced heavily by both Persian and Arabic from the very beginnings of their settling down in Anatolia, given the higher prestige and development of the culture and literature of these neighbouring Muslim nations. The number of works in Turkish written by the Turks of Anatolia (as opposed to those written by them in Arabic and Persian and even Greek) greatly increased in the fourteenth century, together with the Selǰuqi period of feudalism in Anatolia. The gap between the eleventh and thirteenth centuries with respect to the lack of written documents can probably be explained by assuming that the Turkish leaders used Arabic and Persian, not finding a local Turkic language in their new surroundings and not having a strong literary tradition to fall back on – given that these Turkish tribes (to a large extent belonging to the Oγuz) were not among the culturally

more advanced Turkic groups and, moreover, were geographically separated at that time from the Central Asian centres of Turkic literature.

From the very beginning of its Anatolian period, Turkish was written in the Arabic script, until the Latin script was adopted in the course of the so-called 'writing reform' of 1928 (put into force in 1929), one of the various reforms introduced after the founding of the Turkish Republic with the aim of westernising the country. However, the Uighur script was also employed by the Anatolian Turks up to the fifteenth century, which might explain some features of the Arabic script as used by the Turks of that period and which differ from standard Arabic usage, e.g. vowels are written out in Turkish words. This point, incidentally, has often been brought up to motivate the so-called 'writing reform', arguing that the multiple ambiguities that arise in Turkish within a non-vocalised orthography made the Arabic system highly inadequate for Turkish.

The dialect of the earliest Anatolian texts has various features in common with the Oγuz dialect as documented for the eleventh century, before the migration to Anatolia, and with Qïpchaq (an ancient language of the Northwestern group) and Turkmenian. Some of these are listed below:

(1) *d* for *t* in Old Turkic. (A number of these *d*s became devoiced again through assimilation in the fifteenth century.)
(2) Initial *b* changes to *v: bar-* > *var-* 'to go; to arrive'; *ber-* > *ver-* 'to give'
(3) Suffix-initial *γ*, *g* disappears (in most instances; there are some surviving suffixes such as *-gil*, *-gen*, but those aren't productive).
(4) Word-final *γ*, *g* disappears in polysyllabic words.
(5) Instead of the second person plural imperative ending *-ler*, *-lar* in Old Turkic, the forms *-nüz*, *-nuz*, *-niz*, *-nız* are found (and remain until today).

Forms which are limited to Anatolian Turkish are the following:

(1) The suffix *-ecek*, *-acaq* appears for the first time in the thirteenth century (but is used as a participle and not yet as a finite verb, as is also possible in Modern Turkish).
(2) The suffix *-iser*, *-ïsar* is the most widely used suffix for the future tense in Anatolian Turkish between the thirteenth and fifteenth centuries and is seen only very infrequently in some Turkmenian and Qïpchaq works.

However, the differences between Old Turkic and early Anatolian Turkish must not have been great and their phonology essentially identical. The vocabulary is also similar to a large extent, although obviously many borrowings from Islamic sources are seen in the realm of religious-mystical concepts.

In the works of the fourteenth century and afterwards, peculiarities of Eastern Turkic, which had crept into Anatolian Turkish because of the Eastern origins of some authors, disappear almost completely, while the component of Arabic and Persian words and forms increases; such Eastern Turkic features include: initial *m* instead of *b* in words containing a nasal: *men* instead of *ben* 'I'; *min-* instead of *bin-* 'to ride'; initial *b*, which, as mentioned above, changed to *v* in Anatolia and neighbouring areas, remained unchanged in Eastern Turkic and is also sporadically found in early Anatolian works: *ber-* instead of *ver-* 'to give'; and, as another example for a different feature, *bol* instead of *ol* 'to be'.

In the literature written for scholarly, administrative and literary purposes, the Persian and Arabic components became so prevalent that 'Ottoman' became a mixed language,

having lost some of its characteristic Turkic properties to the point of not being usable as a medium of communication common to all social classes. During the same time, however, there also was a considerable production of mystical literature and folk poetry which was written for the less educated classes, in the language used by those segments of the population, namely Anatolian Turkish as influenced very little by Persian and Arabic. These works are very close to the 'Republican Turkish' of today and can essentially be understood without too much difficulty. Among the authors of the 'court literature' there were, time and again, also some who called for a purification of the language and ultimately, starting in the eighteenth century, there was a general movement towards a language with local (rather than foreign) features.

The culmination of such movements was reached after the turn of the century. In 1909, a 'Turkish Club' (*Türk Derneği*) was founded in Istanbul and started publishing a journal, proclaiming its aims for a simpler Turkish. Similar movements and journals followed soon and literary works written in a 'purified' Turkish were produced (see, for instance, the works of Ömer Seyfettin and Ziya Gökalp). Conscious and systematic efforts to establish criteria for maintaining the vocabulary as well as the structural properties of Turkish were continued through the 'War of Liberation' (after World War I) into the founding of the Republic and the reform movements. The language reform, which can be said to have started with the 'writing reform', should therefore be viewed within a tradition of a search for a national identity, combined with a general campaign for westernisation. A Turkish Language Academy was founded in Ankara, with the tasks of etymological research and creation of new words, the latter in accordance with the Turkish rules of word formation and using Turkic roots, where the 'purification' of the language from Arabic and Persian vocabulary had created gaps which could not be filled with current synonyms. Although some of these new creations were judged to be just as foreign to the current colloquial language as the borrowed vocabulary and dropped out of usage almost as soon as they were introduced, the work of the Academy can be viewed as having been essentially successful in creating a widely understood language with a transparent morphological component and its own, typologically consistent syntax.

2 Phonology and Orthography

The vowel inventory of Turkish is very symmetric. The eight phonemic vowels are grouped into foursomes with respect to the features of height, backness and rounding, as in Table 30.1.

All vowels of the native vocabulary are underlyingly (or, say, phonemically) short. There is, however, vocalic length on the surface, having various sources: (1) borrowings with unpredictably long vowels; e.g. *ha:dise* 'event, happening'; *ma:zi:* 'past';

Table 30.1 Turkish Vowels

	[−back] [−*round*]	[+*round*]	[+back] [−*round*]	[+*round*]
[+high]	i	ü	ɨ	u
[−high]	e	ö	a	o

(2) compensatory lengthening in words of Turkic origin, where an original voiced velar fricative (which is no longer part of the *surface* inventory of segments in modern standard Turkish) used to follow a vowel. There are some arguments that show this segment to be part of the *phonemic* inventory, since it behaves like a consonant in stem-final position with respect to allomorphy choice of following suffixes. For example, the accusative and dative suffixes are: *-I* and *-A* after a consonant, but *-yI* and *-yA* after a vowel, respectively. (For the notation with capital letters, see page 526.) After a stem-final phonetically long vowel due to 'compensatory lengthening' (but not after an *inherently* long vowel), the allomorph regularly chosen by consonant-final stems appears; e.g. orthographic *dağ* 'mountain', pronounced *da:*, accusative *daı*, dative *daa*. Compare these forms with: *araba* 'car', accusative *arabayı*, dative *arabaya*, and *bina:* 'building', accusative *bina:yı*, dative *bina:ya.* Where that segment (which is, as shown in the examples above, rendered by the sign *ğ* in Turkish orthography and can never occur word-initially – see also the section on the historical background of Turkish) is in either word-final or pre-consonantal (i.e. in syllable-final) position, the preceding vowel is lengthened; e.g. orthographic *çağ* 'era', pronounced *ča:*, locative *çağda*, pronounced *ča:da*. Note that length doesn't necessarily lead to word-level stress; thus, the last example is stressed on its final, short, vowel, while the long vowel remains without stress: [*ča:dá*]. The same is true when vowel length is due to inherent length in borrowings, rather than to compensatory lengthening: [*bina:yá*].

Another peculiarity of Turkish vowels is that non-high vowels cannot be round, unless they are in a word-initial syllable. While many borrowed stems are exceptional in this respect (e.g. *dekor* 'stage design'; *pilot* 'pilot'), there is only one affix that is exceptional: the progressive suffix *-(I)yor*.

Perhaps the most prominent property of the Turkish vowels is the fact that they undergo vowel harmony with respect to backness and rounding. We shall discuss this issue in more detail later on, when the phonological rule system of the language is investigated.

The consonant inventory of Turkish is given in Table 30.2. The consonants *k*, *g* and *l* have two forms: palatal and velar. Their distribution is, in general, determined by the backness versus frontness of the tautosyllabic vowel, e.g. *čök̟* 'collapse' versus *čok* 'many; very'; *bel̟* 'waist' versus *bal* 'honey'; *k̟ör* 'blind' versus *kor* 'ember'; *il̟ik̟* 'marrow' versus *ılık* 'luke-warm'. (x̟ denotes a palatal consonant.) These assimilative changes are not always predictable, however; there are some borrowings where the

Table 30.2 Turkish Consonants

		Bilabial	*Labio-dental*	*Dental, Alveolar*	*Palato-alveolar*	*Palatal*	*Velar*	*Glottal*
Stop	voiceless	p		t	č		k	
	voiced	b		d	ǰ		g	
Fricative	voiceless		f	s	š			
	voiced		v	z	ž			
Nasal		m		n				
Lateral approximant				l				
Central approximant				r		y		h

palatal variant precedes or follows a tautosyllabic back vowel; e.g. *kalp* 'heart'; *k̲ar* 'profit'. (Where a palatal velar stop precedes a back vowel, a palatal glide is inserted after the palatal velar stop; cf. the section on orthography further below.)

The Latin alphabet used for modern standard Turkish is, both in its printed and handwritten versions, the familiar system used in more familiar European languages – as, for example, in English. The diacritics used for less common sounds make some of the signs very similar to some versions of the phonetic script; for instance, the phonetic symbols for vowels given in Table 30.1 are also the ones used in Turkish orthography, with one exception: Instead of *ɨ*, the sign used for the high back non-round vowel, we find *ı*, i.e. a dotless *i*. The difference between the two non-round high vowels is signalled in the same way for capital letters: *İ* for the front, *I* for the back, high non-round vowel. As for the consonants, we have commented on the 'silent' *ğ* earlier. Other letters that don't correspond to the familiar phonetic symbols are the following: *c* for [ǰ], *ç* for [č], *ş* for [š], *j* for [ž].

The orthographic conventions correspond roughly to those of a broad phonetic transcription. Predictable alternations (e.g. those due to syllable-final oral stop devoicing, to voicing assimilation or to vowel harmony) are written out, differing in this respect from, say, the German orthography. Other predictable alternations are not signalled, however: since there are no special signs for the palatal versus velar *k*, *g* and *l*, the alternations that these segments undergo remain unexpressed by the orthography. Unpredictable occurrences of the palatal variants of these consonants with back vowels *is* sometimes shown, however, by placing a circumflex on the vowel: *kâr* [k̲yar]. (A front glide is inserted when the consonant in question is a k̲.) Inherent vowel length is not shown by the writing, although it is unpredictable.

With the exception of some learned words and the borrowed vocabulary of native speakers who either have some knowledge of European languages or live in big cities with extensive western influence, Turkish does not allow consonant clusters in initial position. In standard pronunciation (and increasingly also in the orthography) such clusters that enter the language via borrowings are broken up by an epenthesised high vowel, which – in general – harmonises with the following vowel(s) in backness and rounding, e.g. learned *k̲lūp* 'club', colloquial *kulüp*; learned *kral* 'king', colloquial *kɨral*.

Turkish is somewhat more tolerant of syllable-final consonant clusters. Three types of clusters are allowed as a coda: (a) sonorant + obstruent: *k̲ent* 'city', *harf* 'letter'; (b) voiceless fricative + oral stop: *çift* 'couple', *şevk̲* 'fervour' (note that the *v* is pronounced [*f*] and thus fits this characterisation involving a voiceless fricative); (c) *k* + *s*: *raks* 'dance', *boks* 'boxing'. Where a stem has a consonant cluster in syllable-final position that does not fall under any of the permissible sets, again a high vowel is epenthesised which undergoes harmony, e.g. 'forehead' accusative *aln* + *ı*, nominative *alın*; 'nose' accusative *burn* + *u*, nominative *burun*; 'city' accusative *şehr* + *i*, nominative *şehir*; 'time' accusative *vakt* + *i*, nominative *vakit*.

A subcase of syllable-final consonant clusters are geminate consonants. While Turkish does tolerate geminate consonant sequences when their members are heterosyllabic (e.g. *et* + *te* 'meat + loc.'), it does not allow them to occupy syllable-final position. Rather than breaking up such clusters by epenthesis, however, the language has a rule of degemination, e.g. 'feeling' accusative *hiss* + *i*, nominative *his*; 'line' accusative *hatt* + *ı*, nominative *hat*.

In addition to the rules discussed above (i.e. vowel epenthesis, consonant degemination), there are a few other important phonological rules that were mentioned in passing and which will receive further attention below.

Syllable-final oral stop devoicing: similar to the more general obstruent devoicing rule in languages like German and Russian, Turkish has a rule that devoices oral non-continuants (i.e. regular stops as well as affricates) in syllable-final position, e.g. *kitap* 'book', accusative *kitab* + *ı*, locative *kitap* + *ta*; *kireç* 'lime', accusative *kirec* + *i*, locative *kireç* + *te*.

The *k/Ø* alternation: the final *k* of a polysyllabic word is deleted phonetically in intervocalic position, where the preceding vowel is short. This *k/Ø* alternation is orthographically rendered as a *k/ğ* alternation, e.g. *kabak* 'pumpkin', accusative *kabağı* [kabaɨ]; *kabuk* 'crust', accusative *kabuğu* [kabuu]. It is possible to view this phenomenon as a subcase of the voiced/voiceless alternation discussed in the previous section. If it is assumed that the alternating *k*s are derived from underlying *g*s as a result of syllable-final stop devoicing and if a rule of intervocalic fricativisation is posited for the voiced velar stop, the data are essentially covered.

Word-final liquid devoicing: another striking phenomenon somewhat related to stop devoicing is the word-final devoicing of liquids, especially common in the Istanbul dialect and in the speech of educated speakers in the other big cities: *kar̥* 'snow', *bakır̥* 'copper', *k̂e̥l* 'bald'. It should be noted, however, that this is not a completely unified phenomenon; some speakers devoice only the palatal *l*, while other speakers do not make a distinction between the two variants of the lateral. (The *r* is devoiced by all speakers who observe the liquid devoicing rule.) It should also be pointed out that liquid devoicing differs from oral stop devoicing in applying at word boundary rather than at syllable boundary; e.g. while the underlying stem-final *b* devoices in: *kitap* + *lık* 'object designated for books; bookshelf', the stem-final *r* remains voiced in a similar environment: *kar* + *lı* 'with snow; snowy' (and not: **kar̥lı*).

Morpheme-initial voicing assimilation: a morpheme-initial obstruent assimilates in voicing to the preceding segment within the word. This rule has to apply after syllable-final stop devoicing has taken place, e.g. (a) *gemi* + *ci* 'sailor' (cf. *gemi* 'ship'), *iz* + *ci* 'boy-scout' (cf. *iz* 'track, trace'), *bakır* + *cı* 'coppersmith' (cf. *bakır* 'copper'); (b) *kitap* + *çı* (cf. *kitap* 'book', underlyingly /kitab/), *şarap* + *çı* 'wine maker, wine seller' (cf. *şarap* 'wine', underlyingly /şarab/).

Vowel harmony: perhaps the most striking property of Turkish phonology is the fact that the distribution of vowels within a word is governed by vowel harmony, i.e. vowels share the specification for the feature [back] and, if they are high, they also share the specification for [round]: *bülbül* + *ümüz* + *ün* 'nightingale + 1 pl. + gen.', 'belonging to our nightingale'; *bülbül* + *ler* + *imiz* + *in* 'nightingale + pl. + 1 pl. + gen.', 'belonging to our nightingales'; *kol* + *umuz* + *un* 'arm + 1 pl. + gen.', 'belonging to our arm'; *kol* + *lar* + *ımız* + *ın* 'arm + pl. + 1 pl. + gen.', 'belonging to our arms'. Note that the [–high] vowel of the plural morpheme, while undergoing vowel harmony for backness, does not undergo rounding harmony. Moreover, since there is a condition (mentioned earlier in this section) on [–high] vowels to the effect that they have to be [–round] if they are in a non-initial syllable, the negative specification of this vowel for rounding is fully determined. Note also that once a non-round vowel follows a round vowel (as in the second and fourth examples above) all vowels to the right of that non-round vowel will be non-round as well, irrespective of their height.

This situation can be characterised in more general terms: where a vowel does not share the specification for a harmony feature with preceding vowels, it will create its own harmony domain, in the sense that it will determine the specification with respect to that particular feature for the following vowels. This description also characterises

the application of vowel harmony where an exceptional vowel occurs. As mentioned before, many stems have exceptional vowels that violate either backness or rounding harmony or both at once; the second vowel of the progressive marker *-(I)yor* is also exceptional in this respect and never alternates. (Capital letters denote archiphonemes whose missing feature values are predictable by rule. In the case of vowels, *I* stands for a [+high], *A* for a [–high] vowel before application of vowel harmony. In the case of consonants, a capital letter stands for a segment which may undergo syllable-final stop devoicing, morpheme-initial voicing assimilation or intervocalic *k*-deletion. Symbols in parentheses denote affix allomorphy in those instances where the segment in question deletes after a 'like' segment (i.e. a vowel after a vowel, a consonant after a consonant).) In such cases, it is the exceptional vowel (or, if there is more than one, the last exceptional vowel) that determines what kind of vowel harmony the following vowels will undergo. Observe the following examples: *dekor* + *un* + *u* 'stage design + 3 sg. + acc.', 'his stage design, acc.', *otobüs* + *ün* + *ü* 'bus + 3 sg. + acc.', 'his bus, acc.'; *buket* + *in* + *i* 'bouquet + 3 sg. + acc.', 'his bouquet, acc.'; *Macar* + *istan* + *ın* + *ı* 'Hungarian + country + 3 sg. + acc.', 'his Hungary, acc.'.

Sometimes, however, a consonant rather than a vowel can determine (backness) harmony. This happens when a palatal consonant unpredictably follows a back vowel in the same syllable and where that consonant is in stem-final position (or a member of a stem-final consonant cluster). In such cases the following vowels will exhibit *front* harmony; i.e. the 'trigger' of vowel harmony will be the exceptional consonant rather than the regular vowel, e.g. *petrol̂* 'petrol, gasoline', accusative *petrol̂* + *ü*; *kal̂p* 'heart', accusative *kal̂b* + *i*; *val̂s* 'waltz', accusative *val̂s* + *i*.

Labial attraction: there are a number of stems with a vowel sequence of *a ... u* and an intervening labial consonant (the latter can also be part of a consonant cluster). Since the second vowel, being high, should undergo rounding harmony, it should surface as an *ı*. Its rounding has traditionally been ascribed to the preceding labial consonant. Some examples are *karpuz* 'watermelon', *kavun* 'melon'. The status of this observation in terms of a rule (of assimilation) in modern standard Turkish has been challenged more recently. While such an assimilatory process might have been productive in Early Anatolian Turkish (and could even have been a feature common to the Southwestern Turkic group), it seems that it is less general in the contemporary language; there are a number of examples where the sequence *a ... ı* shows up in spite of an intervening labial consonant, e.g. *çarmıh* 'cross', *sabır* 'patience', *kapı* 'door' (but note that, interestingly, this item is pronounced as [*kapu*] in some dialects). Furthermore, an even larger number of stems exhibit *a ... u* sequences without any intervening labial consonant; e.g. *ka:nun* 'law', *arzu* 'desire', *fasulya* 'bean'.

Turkish has in general word-final stress: *kitáp* 'book'; *gör* + *ebil* + *ecek* + *lerin* + *í* 'see + abilit. + fut. + 3 pl. + acc.' 'that they will be able to see'. Some suffixes are exceptional, however, in: (a) rejecting stress when in word-final position: *gör* + *ecék* + *ti* 'see + fut. + past' 'he was going to see'; (b) dividing the word into stress domains where not word-final: *gör* + *é* + *me* + *yecek* + *lerin* + *ì* 'see + abilit. + neg. + fut. + 3 pl. + acc.' 'that they will not be able to see'. Under both circumstances, the vowel preceding the exceptional morpheme receives primary stress. More recent literature has attempted to predict exceptionality of stress by applying syntactic criteria; e.g. *-ti*, the past tense morpheme in *gör* + *ecék* + *ti* 'see + fut. + past' is not inherently exceptional, but only if it follows another tense morpheme, as is the case in this example, where it follows the future suffix and thus functions as a perfective marker. If this morpheme is

analysed as a copular past tense with the preceding future morpheme as a (future) participle marker that 'closes off' a 'small word' domain, then that domain receives regular final stress, and this example stops being an exception. Other exceptional suffixes may be less amenable to such an account, but other syntactically based accounts are possible nonetheless.

A rule that applies within a phrase or a compound to reduce stresses left-to-right is needed independently: *dérs kitab + ì* 'course book + compound marker', 'textbook'. This rule can be used to account for the stress in words like *görémeyeceklerinì* which consist of more than one stress domain and exhibit word-final non-primary stress.

Exceptionality with respect to stress is also exhibited by some unsuffixed stems. Such items do not fall into one clearly and independently defined set. Many (but not all) borrowed stems and almost all place names fall under this group, within which there are subregularities: they are stressed on the antepenultimate syllable, if it is the first non-final closed syllable; otherwise, the penultimate syllable is stressed. Some illustrative examples follow: *İstánbul*, *Ánkara*, *İzmir*, *Zongúldak*, *Adána*, *fasúlya* 'bean', *lokánta* 'restaurant'.

3 Morphology

Turkish morphology is agglutinative and suffixing; there are only very few exceptions to the one-to-one relationship between morpheme and function and only one process that is prefixing rather than suffixing, namely reduplication of the first syllable (with an inserted consonant) in intensifying adjectives and adverbs; e.g. *beyaz* 'white', *bembeyaz* 'completely white'; *çabuk* 'fast', *çarçabuk* 'very fast'. The nature of the inserted consonant is hard to predict fully; the general consensus in the literature is that the phonological features of the consonant are based on a dissimilation process.

In the following, a brief survey will be given of the most productive suffixes and some restrictions will be stated that govern their occurrence and the ordering among those morphemes that can co-occur; later on, specific categories of special interest will be discussed. Inflectional suffixes will be referred to as 'verbal' or 'nominal' according to the category of the stem they attach to. By 'nominal stems' are meant nouns, adjectives and adverbs. (Participials and gerundives will fall under the 'nominal' group in this respect.)

As to be expected, derivational suffixes precede inflectional ones. Not surprisingly, among those morphemes that derive nominals, those that attach to verbal stems precede those that attach to nominal ones, where the two types co-occur: *ver* + *im* 'give + abstr. n.', 'profit'; *ver* + *im* + *li* 'give + abstr. n. + with (adj.)' 'profitable'; *ver* + *im* + *li* + *lik* 'give + abstr. n. + with (adj.) + abstr. n.', 'profitability'. The suffixes exemplified in the last two examples can attach to underived nominals, as well: *balkon* + *lu* 'with a balcony; balconied'; *dürüst* + *lük* 'honest + abstr. n.', 'honesty'. Both groups are productive; two other productive members of the first group are the action/manner suffix *-(y)Iş*, the result/action morpheme *-mA* and the infinitive marker *-mAK*. In the second group, we find *-CI*, deriving nouns meaning 'professional', and *-sIz*, deriving adjectives meaning 'without'.

The first member of a sequence of nominal inflectional suffixes and hence immediately following derivational morphemes, if present, is the plural marker *-lAr: gül* + *üş* + *ler* 'laugh + act. n + pl.', 'laughters; manners of laughing'; *at* + *lar* 'horse + pl.', 'horses'.

Next come nominal agreement suffixes. These are often referred to as 'possessive suffixes' in traditional literature, the reason being that the nominal stem they attach to is often, if not always, interpreted as possessed by a noun phrase within the clause or phrase. The reason they are referred to as 'agreement suffixes' here is that they express the person and number features of their 'possessors'. A more detailed account of these suffixes will be offered in the next part of this section which will be devoted to issues of special interest.

Case morphemes occur last, e.g. *üstün + lüğ + ümüz + ü* 'superior + abstr. n. + 1 pl. + acc.' 'our supremacy (accusative)'. The group of agreement morphemes will be discussed in more detail in the second part of this section. It should be mentioned here, however, that not more than one case morpheme can occur within an immediate sequence of suffixes.

There is only one completely productive morpheme that derives verbs from nominals: *-lA*, which has a meaning related to the causative; e.g. *karşı + la + mak* 'opposite + deriv. morph. + infin.', 'to go to meet; to respond; to reply to'; *kara + la + mak* 'black + deriv. morph. + infin.', 'to blacken'. This morpheme can then be followed by the various verbal suffixes which we shall briefly discuss according to the sequential order in which they occur within the word.

The leftmost productive class in the string of verbal suffixes is the category often called 'voice' by traditional grammars. This group consists of the middle/reflexive (*-(I)n*), the reciprocal (*-(I)ş*), the passive *(-Il/n)* and the causative *(-DIr/t)*. (The *-Il* allomorph of the passive follows consonants, the *-n* allomorph follows vowels. *-DIr* is the basic allomorph of the causative; *-t* occurs after polysyllabic stems which end in a vowel or in the oral sonorants *r* and *l*.) The middle/reflexive and the reciprocal cannot co-occur; where the passive co-occurs with either one, it has to follow them. In the very few examples where the causative can co-occur with the middle/reflexive and the reciprocal it has to follow them, and, while it can co-occur with the passive, it has to precede it; e.g. *tanı + ş + tır + ıl + dı + lar* 'know + recip. + caus. + pass. + past + 3 pl.' 'they were caused to know each other; they were introduced to each other'.

Suffixes of this group can be followed by the verbal negation marker *-mA*, which is one of the suffixes that are exceptional from the point of view of word stress in rejecting word-final stress and causing the preceding vowel to be stressed. This suffix, in turn, is followed either by one of the various mood markers or by purely verbal or gerundive/participial forms, the latter expressing tense in varying degrees of differentiation. The mood markers are: the desiderative *-sA*, the necessitative *-mAlI* and the optative *-(y)a;* e.g. *gör + üş + me + meli + yiz* 'see + recip. + neg. + necess. + 1 pl.', 'we shouldn't/mustn't see each other'. The suffixes of the mood category are mutually exclusive.

The tenses are: definite past: *-DI*; reported past: *-mIş*; aorist: *-(A)r*; future: *-(y)AcAK*; present progressive: *-(I)yor*. These forms have also aspectual connotations: the past tenses denote accomplished actions and the aorist actions that are either extended or repeated over a period of time. The present progressive is similar to its English equivalent in denoting an action that, roughly speaking, takes place at the time of the utterance. One difference is that stative verbs, unlike those in standard English, can take the progressive in Turkish:

ev	+	e	git +	mek	isti +	*yor*	+	um
home	+	dat.	go +	infin.	want +	pres. prog.	+	1 sg.

'I want (*am wanting) to go home'

The main participial forms are those used in relative clauses: *-(y)An* and *-DIK*, and they will be discussed in Section 4. Also in this group (from the point of view of positional slots within the morphological word) are so-called verbal nouns and converbs (these are terms often used in traditional literature). The 'verbal nouns' consist mainly of the infinitive suffix *-mAK* and the result/action noun marker *-mA* and were also listed among the derivational morphemes that convert verbs into nominals. *-DIK* is a general factive nominaliser that corresponds rather closely to the English gerundive *-ing*. It has non-future temporal value; for future, the suffix *-(y)AcAK* is used, which doubles as a finite future marker. Converbs (or gerundives, as they are also called) are suffixes that yield adverbial forms. Some examples are the manner suffix *-(y)ArAk*, the conjunction adverbial *-(y)Ip* which denotes close successions of actions and the time adverb suffix *-(y)IncA*. In general, only one of the suffixes in this group can occur at a time. In other words, within the morphological sequence, the various gerundive, participial and nominal markers take the place of the tense or mood markers, whether they have tense connotations themselves or not.

However, two tense markers (as well as a tense and a mood marker) *can* co-occur in immediate succession to form complex tenses; in such examples, it might be appropriate to view the second marker as a copula carrying the main tense or mood and the preceding sequence as a participial:

imtihan	+	ım	+	a	başlı	+	*yor*	+	*du*	+ m
exam	+	1 sg.	+	dat.	start	+	prog.	+	past	+ 1 sg.

'I was starting my exam (when…)'

Note that in such sequences, the present progressive marker *-(I)yor* retains its aspectual meaning.

The reported past marker *-mIş* is used as a perfective aspect marker in such sequences (i.e. when it is the first member of the sequence):

imtihan	+	ım	+	a	başla	+	*mış*	+	*tı*	+ m
exam	+	1 sg.	+	dat.	start	+	perf.	+	past	+ 1 sg.

'I had started my exam (when…)'

All tense and some mood markers can occur as the first members in these sequences; however, only the two past tense markers and the mood marker for the desiderative (the latter as a conditional) can occur as the second member, i.e. as the main tense or modality marker. However, all the tenses can be used as a main tense or modality within a periphrastic construction with an auxiliary verb. The most widely used auxiliary is the verb *ol-* 'be, become'; e.g.

imtihan + ım	+	a	başlı	+	*yor*	ol	+	*acağ*	+	ım
exam + 1.sg.	+	dat.	start	+	prog.	be	+	fut.	+	1 sg.

'I shall be starting my exam…'

imtihan + ım	+	a	başla	+	*mış*	ol	+	*acağ*	+	ım
exam + 1.sg.	+	dat.	start	+	perf.	be		fut.		1.sg.

'I shall have started my exam'

This mixed positional group is followed by agreement markers, wherever such markers are possible. (Among the suffixes that cannot be followed by agreement markers are the infinitive marker *-mAK*, the participial marker *-(y)An* (unless it functions as a verbal noun) and the gerundive marker *-(y)Ip*.)

Now that we have looked at the most productive morphemes and some regularities of their distribution, let us discuss some typological characteristics of the morphological system.

Gender is neither overtly expressed in nouns (or pronouns), nor does it affect agreement. Agreement itself (by which term we shall mean agreement of the verbal or nominal head of a construction with its subject (or possessor) in terms of the features of person and number) can be either verbal or nominal; in other words, there are two basic slightly different paradigms, given in Table 30.3. (Additional, partly defective, paradigms are not included here, due to their more peripheral status.)

The verbal paradigm appears with the predicates of main clauses and of 'direct complements' (for discussion of the latter, see Section 4); the nominal paradigm is used on the head nouns of possessive noun phrases as well as on the nominalised verbs of gerundive and participial complements. Some illustrative examples follow:

Verbal agreement used with a main clause predicate verb:

(Ben)	bu	makale + yi	yarın	bitir + eceğ + *im*
I	this	article + acc.	tomorrow	finish + fut. + 1 sg.

'I shall finish this article tomorrow'

(Biz)	her	akşam	çok	çalış +	ır + *ız*
we	every	evening	a lot	work +	aor. + 1 pl.

'We work a lot every evening'

Verbal agreement used with a main clause predicate adjective:

(Ben)	bugün	çok	yorgun + *um*
I	today	very	tired + 1 sg.

'I am very tired today'

(Siz)	çok	güzel + *siniz*
you	very	pretty + *2pl.*

'You are very pretty'

Table 30.3 Agreement Markers

	Verbal	Nominal
1 sg.	-Im	-(I)m
2 sg.	-sIn	-(I)n
3 sg.	-∅	-(s)I(n)
1 pl.	-Iz	-(I)mIz
2 pl.	-sInIz	-(I)nIz
3 pl.	-lAr	-lArI(n)

Notes: As before, the suffix-initial vowels in parentheses are deleted after a stem-final vowel; the suffix-initial consonant in parentheses is deleted after a stem-final consonant. The suffix-final consonant in parentheses is deleted in word-final position.

Nominal agreement in a possessive noun phrase:

(Biz-im) heykel + *imiz*
we-gen. statue + *1pl.*
'our statue'

Ayşe-nin araba + *sı*
Ayşe-gen. car + *3sg.*
'Ayşe's car'

Nominal agreement used in a gerundive complement:

Herkes	[(biz + im)	heykel + i	kır + dığ +	*ımız*] + ı	
everybody	we + gen.	statue + acc.	break + fact.n. +	*1pl.* +acc.	
		bil	+ iyor		
		know	+ pres.prog.		

'Everybody knows that we broke the statue'

Herkes	[Ayşe + nin	heykel + i	kır + ma	+ *sın*] + ı	isti + yor
Everybody	Ayşe + gen.	statue + acc.	break + act.n.	+ *3.sg.* +acc.	want + pres.prog

'Everybody wants Ayşe to break the statue'

Another property of Turkish agreement worth remarking on is the lack of it where modifiers are concerned. This means that neither singular/plural properties of a noun nor its case marking will 'spread' onto its adjective modifier(s) or any of its determiners. As a matter of fact, another striking property of Turkish in this respect is the lack of overt plural marking on a noun where its quantifier clearly expresses plurality; this generalisation holds irrespective of the grammatical relation of the noun phrase involved. The following examples will illustrate this point:

Subject noun phrase:

Beş	adam	(*adam + *lar*)	heykel + i	kır + dı
five	man	(man + pl.)	statue + acc.	break + past

'Five men broke the statue'

Indirect object noun phrase:

Beş	adam + a	(*adam + *lar* + a)	yardım	et + ti + m
five	man + dat.	(man + pl. + dat.)	help	do + past + 1 sg.

'I helped five men'

Let us now return to subject-head agreement. The two paradigms in Table 30.3 might be slightly misleading in that the suffixes for plural subjects are presented as unanalysed morphemes. However, especially the nominal paradigm in Table 30.3 can substantiate a possible claim that, at least for the first and second person plural forms, those suffixes consist of two morphemes: 1 sg. *-(I)m*, 2 sg. *-(I)n*, 1 pl. *-(I)m + Iz*, 2 pl. *-(I)n + Iz*. Hence, it would make sense to view the suffix *-Iz* as a plural marker. (This plurality would have to be confined to subject agreement, however, since the general plurality morpheme, *-lAr*, is different.)

The same analysis carries over to the verbal paradigm, if it is assumed that the suffix for person is, idiosyncratically, unrealised in the first person plural agreement form: 1 sg. *-Im*, 2 sg. *-sIn*, 1 pl. *-Ø + Iz*, 2 pl. *sIn + Iz*. The agreement suffixes for third person plural subjects do not seem to fall under this generalisation, simply because their shape is rather different from those of the first and second person plural agreement morphemes. However, we would like to claim that there, too, a further analysis into a person morpheme, distinct from a number morpheme, is possible. Once again, we shall start with the nominal paradigm, which is more perspicuous than the verbal paradigm, since all morphemes are overtly realised: 3 sg. *-(s)I(n)*, 3 pl. *-lAr + I(n)*. Two factors are worth noticing: in comparison with the agreement forms for first and second person plural subjects, the order between the person and number suffixes is switched around, i.e. the number morpheme precedes the person morpheme. In addition, the number morpheme itself is suppletive. Instead of the form *-Iz*, the agreement morpheme for (plural) number exhibited elsewhere in both paradigms, we see here the general plurality morpheme *-lAr*. (Note, incidentally, that the suffix for third person appears in a perfectly regular shape: we know that the parenthesised initial *s* of that morpheme is deleted after a consonant. Since, within the third person plural agreement form, the third person suffix always follows the plural number suffix – and hence an *r* – that suffix will always surface without that *s*.)

Once again, the analysis carries over to the verbal paradigm. The agreement form for third person plural exhibits the suppletive morpheme *-lAr* for plural number. Since the third person morpheme remains unexpressed in the verbal paradigm, nothing else but the plural *-lAr* is included in the total form of the third person plural morpheme, as the last line of Table 30.3 shows.

Yet another peculiarity of the third person plural morpheme is that, under some circumstances, it can be omitted. Essentially, when the subject noun phrase is overtly present (as we shall see in Section 4, subjects can be omitted), the plural 'submorpheme' is optional (and, as a matter of fact, its omission is stylistically preferred):

Adam + lar	heykel + i	kır + dı (+ *lar*)
man + pl.	statue + acc.	break + past *(+3pl.*)

'The men broke the statue'

Hasan [adam	+ lar + ın	heykel + i	kır + dık + *lar* + *ın* + ı/
Hasan man	+ pl. + gen.	statue + acc.	break + fact.n. + *pl.* +*3pers.* + acc./
		kır + dığ + *ın* + ı]	bil + iyor
		break + fact.n. + *3pers.* + acc.	know + pres. prog.

'Hasan knows that the men broke the statue'

None of the other agreement morphemes exhibits this freedom of partial occurrence.

Yet another property that determines the occurrence of the plural 'submorpheme' of third person plural agreement is the animacy of the subject noun phrase. The stylistic preference we mentioned in favour of omitting the morpheme in question strengthens to the point of almost a grammatical prohibition against its occurrence when the subject is inanimate:

Kitap + lar	masa + dan yer + e	düş + tü	(??/* + *ler*)
book + pl.	table + abl. floor + dat.	fall + past	(*3pl.*)

'The books fell from the table to the floor'

Let us now turn to the case system in Turkish. It is a matter of some controversy how many cases Turkish has. Traditional Turkish grammars usually assume five cases: nominative: not marked overtly; accusative: *-(y)I*; dative: *-(y)A*; locative: *-DA*; ablative: *-DAn*. It is a fairly well-known phenomenon that the two 'structural' (i.e. non-oblique) cases, the accusative and the genitive, don't show up when the constituent in question (i.e. the direct object or the subject of a nominalised clause, respectively) is non-specific (non-referential):

Ali kitab + ı oku + du
Ali book + *acc.* read + past
'Ali read the book'

Ali bir kitab + ı oku + du
Ali a book + *acc* read + past
'Ali read a (particular, specific) book'

Ali bir kitap oku + du
Ali a book read + past
'Ali read a (non-specific) book'

Ali kitap oku + du
Ali book read + past
'Ali read books' ['Ali book-read']

[Ali + yi bir arı + *nın* sok + tuğ + un] + u duy + du + m
Ali + acc. a bee + *gen.* sting + fact.n. + 3.sg. + acc. hear + past + 1.sg.
'I heard that a (particular) bee stung Ali'

[Ali + yi (bir) arı sok + tuğ + un] + u duy + du + m
Ali + acc. (a) bee sting + fact.n. + 3.sg. + acc. hear + past + 1.sg.
'I heard that (a) bee (non-specific) stung Ali' ['I heard that Ali got bee-stung']

The non-specific constituents that lack morphological case marking are quite rigidly limited to immediate pre-verbal position – a rather striking observation, given the otherwise rather free word order in Turkish (to be addressed in the next section on syntax). In this context, some of the relevant literature has analysed such non-specific constituents lacking structural case morphemes as 'incorporated' into the verb. It has become a matter of controversy in some literature what the nature of this incorporation is. The morphological incorporation found in polysynthetic languages is clearly different from the Turkish type, as in Turkish, the verb doesn't change its valency. I will not take a stand here on this matter.

Turkish has also a genitive: *-(n)I(n)*, and an instrumental: *-(y)lA*. It is probably because the genitive is not 'governed' by verbs, but is rather a structural property of the subjects of nominal phrases or clauses, that many grammarians were reluctant to recognise it as a regular case. As for the instrumental, it is a cliticised form of a formerly unbound morpheme; from the synchronic point of view, there are two criteria that could argue against viewing it as a case morpheme: (a) it is exceptional from the point of view of stress (as are all other cliticised morphemes), while all other case morphemes (including the genitive) are regular; (b) it follows the genitive when it is suffixed to a personal

pronoun and hence behaves like a postposition that governs a case – namely the genitive in this instance – and not like a regular case morpheme, which can never immediately follow another case suffix, as was mentioned earlier. We shall not take a stand here on this issue.

In conjunction with the discussion about the status of the instrumental, it should be mentioned that various postpositions 'govern' certain cases, similarly to verbs. The point of interest within the context of morphology is that regular nouns are treated differently from pronouns in this respect by those postpositions that take objects in the nominative. Specifically, while full nouns appear in the nominative in those contexts, pronouns have to be marked with the genitive: *kadın gibi* 'like a woman', *Ahmet kadar* 'as much as Ahmet'; but: *ben* + *im gibi* 'I + gen. like', 'like me', *ben* + *im kadar* 'I + gen. as much as', 'as much as I'.

4 Syntax

Turkish is a perfect example of a left-branching type of language where governed elements precede their governors, i.e. objects precede the verb, the postpositional object precedes the postposition and the (adjective, genitive, or quantificational etc.) modifier precedes the modified head.

The unmarked word order in sentences is SOV; if there is more than one object, and if one of them is a direct object, the order with the indirect object closer to the verb seems less marked than others, at least with certain verbs:

Hasan	kitab + ı	çocuğ + a	oku + du
	book + acc.	child + dat	read + past

'Hasan read the book to the child'

However, other orders are possible, as well. As a matter of fact, Turkish is rather free in its word order. Often (but not always), the divergences from the unmarked order have a pragmatic, discourse-oriented function, in that the position immediately preceding the verb is the preferred focus position and the sentence-initial position is the preferred topic position. New information and material stressed for emphasis appear in focus position and, in addition to being syntactically marked in this way, also receive intonational stress. The topic, i.e. the material that the sentence is about, is placed at the beginning of the sentence and is often separated from it – orthographically by a comma and by a slight pause in speech.

Differently from other SOV languages (e.g. Japanese), Turkish is so lenient about non-canonical word orders that it even permits non-verb-final constructions. Such sentences arise when material is added as an afterthought or when the speaker assumes the hearer to know about it: *Hasan çocuğ* + *a ver* + *di elma* + *yı* 'Hasan child + dat. give + past apple + acc.', 'Hasan gave the apple to the child'. For this example to be felicitous, it must be clear within the discourse that something happened to the apple or even that Hasan gave the apple to somebody. Some recent literature claims that such post-verbal, backgrounded constituents have the same pragmatic as well as structural properties as topicalised constituents. This claim has been successfully contested on structural grounds, showing that post-verbal constituents have different syntactic behaviour from pre-verbal ones. On the other hand, the pragmatic aspect of this issue is still being debated in relevant literature.

An embedded sentence takes up the same position that the corresponding noun phrase with the same grammatical relation would and can move around within the main clause with the same ease as a regular noun phrase:

Hasan ban + a	[imtihan + *ı*	*geç* + tiğ + in] + *i*	anlat + tı
Hasan I + dat.	exam + acc.	pass + fact. nom. + 2 sg. + acc.	tell + past

'Hasan told me that you passed the exam'

Hasan ban + a	anlat + tı	[imtihan + ı geç + tiğ + in] + i

The constituents of the embedded sentence are somewhat less free in their word order. While they can still successfully violate the canonical SOV order within their own clause, they have to move to the right of the highest sentence when they cross the boundary of their own clause and cannot 'scramble into' higher material; thus, compare the following examples with the last set of examples above:

Ahmet ban + a	[— geç + tiğ + in] + i	anlat + tı imtihan + ı
*Ahmet ban + a	[— geç + tiğ + in] + i	imtihan + ı anlat + tı

(The original site of the 'scrambled' constituent is marked with a —.)

In possessive noun phrases, the possessor precedes the head noun; in 'regular' noun phrases, modifiers precede the head. Where there is both an adjectival modifier and an article (only the indefinite article is overtly expressed in Turkish), the adjective precedes the article; where there is both a numeral and an adjective, the unmarked order is for the numeral to precede the adjective:

Ahmed	+ in	kitab + ı	
Ahmet	+ gen.	book + 3 sg.	'Ahmet's book'
ilgi	+ nç	bir kitap	
interest	+ ing	a book	'an interesting book'
üç	ilginç	kitap	
three	interesting	book	'three interesting books'

The genitive-marked possessor can 'scramble' in either direction, while the article and numerals cannot. The adjective is not free to move, either, as far as spoken language and written prose are concerned. In poetry, however, an adjective can occur to the right of its head. Let us also mention, without going into details, that parts of nominal compounds cannot scramble and that postpositions cannot be stranded.

One striking characteristic of Turkish is that a subject can be left unexpressed in finite clauses (i.e. those exhibiting some type of subject-predicate agreement) as well as in possessive noun phrases:

— okul + a	gid +	eceğ + im
school + datgo +		fut. + 1 sg.

'I shall go to school'

Ahmet	[— kitab + ım] + ı	kayb + et + miş
	book + 1 sg. + acc.	loss + do + rep. past

'It is said that Ahmet lost my book'

(The sites of the missing constituents are marked with a —.) This possibility has been traditionally linked to the rich agreement morphology of Turkish, i.e. to the fact that agreement suffixes will uniquely 'identify' the person and number of the subject which is unexpressed.

Although Turkish has no agreement markers for non-subjects, it is also possible to 'drop' such constituents; e.g. *bul* + *du* + *m* 'find + past + 1 sg.' 'I found (it)'. Such examples are more restricted, however, than 'subject-drop' examples. They can never start a discourse, while 'subjectless' finite sentences can. Such constructions are felicitous only if the antecedent of the 'dropped' constituent has been mentioned in the discourse or has somehow been made clear by a pragmatic act.

As we saw before, passive is marked by the morpheme *-Il* (with a morphophonemic alternant *-n*) on the verbal stem. From the syntactic point of view, there are two types of passive constructions; they will be referred to as 'transitive passive' and 'intransitive passive'; the former type is derived from transitive verbs, the latter from intransitive ones. By 'transitive verb' we mean verbs that take direct objects (noun phrases that are marked accusative when they are specific), and by 'intransitive verbs' we mean verbs that do not take such objects (i.e. that either lack objects altogether or take only indirect or oblique objects). The two constructions exhibit the following surface differences: the patient of the action (or, in other words, the direct object of the corresponding active sentence) is the subject of the transitive passive construction. This claim is substantiated by the fact that these subjects exhibit syntactic properties typical of subjects in general: they appear in the nominative case, and in sentence-initial position when the word order is unmarked; the verb agrees with these constituents, as it does with subjects in general; these constituents can show up as the accusative-marked subjects of clauses that act as complements to 'believe-type' verbs (see page 540); they can correspond to the understood subjects of infinitivals; and the agent of the action can appear in an agentive phrase. However, the non-accusative objects that an intransitive verb might co-occur with are not surface subjects in an intransitive passive construction (in the sense that they do not exhibit the criteria just enumerated), and agentive phrases are judged to be awkward at best, if not completely ungrammatical. Some illustrative examples for these differences follow:

(*Biz*)	döv + *ül* + dü + *k*
we	hit + pass. + past + 1 pl.
'We were hit'	

Biz + *e*	yardım	ed + *il* + di
we + dat.	help	do + pass. + past
'We were helped'		

*biz(+ e)	yardım ed + il + di + *k*
We (+ dat.)	help do + pass. + past + *1.pl.*
Intended reading: 'We were helped'	

$(Biz)_i$	$[PRO_i$	döv + ül + mek]	iste + mi + yor + uz
we		hit + pass. + infin.	want + neg. + pres. prog. + 1 pl.
'We don't want to be hit'			
*(biz_i)	$[PRO_i$	yardım	ed + il + mek] iste + mi + yor + uz
		help	do + pass. + infin. want + neg. + pres. prog. + 1 pl
Intended reading: 'We don't want to be helped'			

Obviously, verbs that do not take any objects at all can also appear in impersonal passive constructions:

Koş + ul + du
run + pass. + past
'It was run (i.e. running took place)'

Eğlen + il + di
amuse + pass. + past
'Fun was had'

Agentive phrases are ungrammatical in such objectless constructions.

It has been claimed in some relevant literature that only verbs with agentive semantics can enter the intransitive passive construction, but that stative verbs cannot. While this generalisation does hold for most cases in Turkish, it is possible to find examples where non-agentive verbs can successfully enter the construction. Such examples are best when combined with a 'tense' that has an aspectual connotation of duration (rather than, say, momentary or completed action):

Böyle bir hava + da iyi uyu + n + ur
such a weather + loc. good sleep + pass. + aor.
'One sleeps well in such a weather'

Compare this quite acceptable sentence with the following ungrammatical ones:

*şimdi iyi uyu + n + uyor
now good sleep + pass. + pres. prog.
'Now it is being slept well (i.e. one is sleeping well now)'

*Dün bütün gün uyu + n + du
yesterday wholeday sleep + pass. + past
'Yesterday it was slept the whole day (i.e. one slept the whole day yesterday)'.

Turkish has various wh-question particles most of which are morphologically derived from the particle *ne* 'what': *ne* 'what', *neden* 'why', *niçin* 'why', *niye* 'why', *hangi* 'which', *kim* 'who'. These elements, which are inherently focused, can be found in various positions, as focused constituents can in general. However, again similarly to focused elements in general, they are preferred when they immediately precede the verb. Some relevant literature calls the latter type of focus 'presentational focus', and focused constituents elsewhere in the sentence 'contrastive focus'.

Çocuğ + a kitab + ı kim ver + di
child + dat. book + acc. who give + past
'Who gave the book to the child?'

Yes–no questions are formed by suffixing the particle *-mI* to the constituent questioned; if the whole sentence is questioned, the particle is attached to the verb, preceding the subject agreement markers in simple tense/aspect forms (with the exception

of the simple past and the conditional, where -mI follows the agreement marker) and preceding the copula and its tense and agreement markers in complex forms:

(Sen) çocuğ + a kitab + ı ver + di + n + mi
you child + dat. book + acc. give + past + 2 sg. + mI
'Did you give the book to the child?'

(Sen) çocuğ + a kitab + ı ver + ecek + mi + y + di + n
you child + dat. book + acc. give + fut. + mI + cop. + past + 2 sg.
'Were you going to give the book to the child?'

It should be noted that the particle *-mI* exhibits dual behaviour with respect to the phonology of the language: it is exceptional from the point of view of word stress (rejects domain-final stress), but regular with respect to vowel harmony.

A few examples follow where *-mI* takes a constituent into its scope:

Çocuğ + a kitab + ı sen + mi ver + di + n
child + dat. book + acc. you + mI give + past + 2 sg.
'Was it you who gave the book to the child?'

(Sen) kitab + ı çocuğa + a + mı ver + di + n
'Was it the child that you gave the book to?'

(Sen) çocuğ + a kitab + ı + mı ver + di + n
'Was it the book that you gave to the child?'

The translations show that such constituent questions correspond to clefted questions in English. (Turkish has also a cleft construction which can enter yes–no questions; formally, the construction consists of a relative clause lacking a head noun.) Note that the questioned constituent is located in the (presentational – see above) focus position.

One general property of embedded sentences in Turkish is that they lack complementisers that introduce (or terminate) clauses, as say the complementisers *that* or *for ... to* in English. But a perhaps even more striking characteristic feature of such clauses is exhibited by their predicates: rather than being fully finite in exhibiting the various tense and aspect markers and their combinations as is the case with verbs of main clauses, the predicates of embedded clauses are 'nominalised' with the help of various morphemes (as we saw in the section on morphology). We also saw that the subject agreement markers on these 'nominalised' predicates come from the nominal rather than the verbal paradigm; one additional criterion for calling these clauses 'nominalised' is that their predicates carry overt case markers:

[Ahmed + in ben + i sev + diğ + in] + *i*
Ahmet + gen. I + acc. love + fact. nom. + 3 sg. (nom.) + *acc.*
bil + iyor + um
know + pres. prog. + 1 sg.
'I know that Ahmet loves me'

[Ahmed + in ben + i sev + me + sin] + *i* isti + yor + um
Ahmet + gen. I + acc. love + act.n. + 3 sg. (nom) + *acc.* want + pres.prog. + 1.sg.
'I want Ahmet to love me'

The two 'nominalisation' morphemes (i.e. *-DIK*, the 'factive nominalisation', and *-mA*, the non-factive 'action nominalisation') exhibited above are the forms exhibited by embedded clauses that function as arguments of the verbs of the higher clause. The semantics of that higher verb and the propositional properties of the clause determine which one of the two morphemes will be chosen, as illustrated by the examples above and their translations.

A subset of the verbs that take clauses with the 'action' (or non-factive) nominal marker also take clauses that are marked with the infinitive suffix *-mAK*. These are comparable to English infinitivals in that they necessarily lack overt subjects; note also that they do not carry agreement morphology:

Ben [sev + il + mek] isti + yor + um
I love + pass. + infin. want + pres. prog. + 1 sg.
'I want to be loved'

Infinitivals can take case markers, too, and are thus shown to be genuine nominalised clauses, as well:

Ahmet [ben + i sev + meğ] + *e* başla + dı
Ahmet I + acc. love + infin. + *dat.* start + past
'Ahmet has started loving me'

Clauses which are postpositional objects and adverbial clauses are also nominalised; in part, their morphology and syntax are similar to those of argument clauses as illustrated above and in part somewhat different. But to discuss these details would go beyond the scope of this chapter. A brief overview follows.

The most typical adverbial clauses are ones where the predicate bears a special, non-finite morphological form that has no morphological subject agreement suffix and which is traditionally called a 'converb'. The subject is typically (and, in manner adverb clauses, obligatorily) omitted; where the subject does show up, it is in the nominative (rather than in the genitive, as might be expected in a non-finite clause):

Ali oda + ya [— gül + *erek*] gir + di
Ali room + dat. laugh + manner conv. enter + past
'Ali entered the room laughingly' (The 'silent' subject is represented with a — space.)

Ali [— yorul + *unca*] ev + e dön + dü
Ali tire + temp. conv. home + dat. return + past
'Ali returned home when (he) got tired' (The 'silent' subject is represented with a — space.)

Ali [Oya yorul + *unca*] ev + e dön + dü
Ali Oya (nom.) tire + temp. conv. home + dat. return + past
'Ali returned home when Oya got tired'

Adverbial clauses can also bear the factive and the non-factive (= action) nominalisation markers:

[[Oya + nın yemek pişir + me + si] için] (ben) ev + de kal + dı + m
Oya + gen. food cook + act.n. + 3.sg. for I home + loc.stay + past + 1.sg.
'I stayed home so that Oya should/could cook' (' … for Oya to cook')

[[Oya yemek pişir + diğ + i] için] (ben) ev + de kal + dı + m
Oya (nom.) food cook + fact.n. + 3.sg. for I home + loc. stay + past + 1.sg.
'I stayed home because Oya cooked'

Note that the subject of the non-factive 'action' adverbial clause is in the genitive, just as it is also in the corresponding argument clauses. However, the subject of the factive adverbial clause is in the nominative, in contrast to the corresponding argument clause, but patterning with 'converb' clauses, as illustrated above. There is a body of recent literature that addresses these properties and contrasts.

A very small subset of embedded clauses exhibits verbal morphology and syntax identical to that of main sentences. Such clauses occur with verbs of belief and are, essentially, interchangeable with corresponding *-DIK* clauses (i.e. factive nominals) which can also be taken by verbs of belief. In some of the few instances where these constructions have been noted, they have been called 'direct complements'. They are of two types:

(a) the embedded subject is marked nominative; the embedded verb exhibits regular verbal subject agreement marking:

Herkes [(ben) üniversite + ye başla + *yacağ* + *ım*] san + ıyor
everybody I (nom.) university + dat. start + fut. + 1 sg. believe + pres. prog.
'Everybody believes that I shall start university'

(b) the embedded subject is marked accusative; the embedded verb exhibits only tense/aspect marking, but no agreement marking:

Herkes [ben + *i* üniversite + ye başla + yacak] san + ıyor
everybody I + acc. university + dat. start + fut. believe + pres. prog.
'Everybody believes me to be starting university'

In addition, there are speakers who also accept a hybrid form where the embedded subject is accusative, but where the embedded verb exhibits regular verbal agreement markers:

Herkes [ben + *i* üniversite + ye başla + yacağ + ım] san + ıyor
everybody. I + acc. university + dat. start + fut. + 1 sg. believe + pres. prog.
(Same translation as the previous example.)

Note that no speakers accept such tensed 'direct' complement clauses, when the embedded verb doesn't exhibit subject agreement marking and when the embedded subject is in the nominative:

*Herkes	[ben	üniversite +	ye	başla +	yacak]	san +	ıyor
everybody	I (nom.)	university +	dat.	start +	fut.	believe +	pres. prog.

Intended reading: 'Everybody believes me to be starting university'

Like all modifiers in the language, relative clauses in Turkish precede their heads. The verbs of such clauses are nominalised, and just as is the case with all regular embedded clauses, they lack complementisers. There is a gap in the position of the constituent within the clause that corresponds to the head.

The factive nominal marker *-DIK* is the basic type of morphology in these constructions; *-mA*, the 'result action' nominal, never occurs, and neither does the infinitive. *-DIK* is replaced by the morpheme *-(y)An* where the 'relativised' constituent is a subject, part of a subject or a non-subject of a clause that lacks a subject (e.g. of an intransitive passive construction as in the last example below); yet another difference between the two constructions follows from this last property: *-DIK* is, as usual, followed by nominal agreement morphology; *-An* never is:

[Ahmed +	*in*	git +	*tiğ* +	*i*]	okul
Ahmet +	gen.	go +	*DIK* +	3 sg.	school

'the school that Ahmet goes to'

[okul +	a	gid +	*en*]	çocuk
school +	dat.	go +	(y)*An*	child

'the child that goes to school'

[[oğl + u]	okul +	a	gid +	en]	adam
son + 3 sg.	school +	dat.	go +	(y)*An*	man

'the man whose son goes to school'

[gid + *il* +	en]	okul
go + pass. +	(y)*An*	school

'the school that is gone to'

Embedded questions have essentially the shape of regular embedded clauses: they are nominalised. Only *-DIK*-clauses can be embedded questions; *-mA*-clauses cannot. (This probably goes together with the fact that *-DIK*-clauses are independent from the main clause with respect to tense and aspect, since they are overtly marked for at least the future/non-future distinction; *-mA*-clauses lack tense completely and are dependent on the main clause for tense and aspect.) This does not mean that *wh*-elements cannot occur within *-mA*-clauses; when they do, however, the main clause is interpreted as a question rather than the embedded clause, while with *-DIK*-clauses either interpretation is possible:

Ahmet	[okul + a	kim + in	git + tiğ + in] + i	duy + du
Ahmet	school + dat.	who + gen.	go + fact.n. + 3 sg. + acc.	hear + past

This has the embedded question reading: 'Ahmet heard who went to school' and the main clause question reading: 'Who did Ahmet hear goes to school?' (i.e. 'about whom

did Ahmet hear whether he goes to school?'). (These two interpretations are distinguished intonationally, with falling intonation on the main clause verb for the former and slightly rising intonation for the latter.) This ambiguity disappears when the question element occurs with a *-mA*-clause:

Ahmet [okul + a kim + in git + me + sin] + i isti + yor
Ahmet school + dat who + gen. go + act.n. + 3 sg. + acc. want + pres. prog.
'Who does Ahmet want to go to school?'

The embedded question reading is not possible: '*Ahmet wants whom to go to school.'

Yes–no questions are also basically similar to regular embedded clauses, particularly where constituents of the embedded clause are questioned; however, where the whole embedded clause is questioned, and where attachment of the question particle *-mI* is expected on the embedded verb, a periphrastic construction in the shape of a participial coordinate structure is found instead (sometimes referred to as an 'A-not-A construction'):

Ahmet [(ben + im) okul + a gid + ip git + me + diğ + im] + i
Ahmet I + gen. school + dat go + and go + neg. + fact.n. + 1 sg. + acc.
sor + du
ask + past
'Ahmet asked whether I go/went to school (or not)'

It is ill-formed to say, with this meaning,

*Ahmet [(ben + im) okul + a git + tiğ + im + i + mi] sor + du,

although this is grammatical with the interpretation 'is it about my going to school that Ahmet asked?'

One more construction with a main/embedded clause asymmetry in the sense that a given constraint holding of the embedded structure does not hold of the main clause is verb-gapping in coordinate structures. In main clause coordinate structures with identical verbs, either the first or the second conjunct can lack its verb:

Ahmet balığ + ı pişir + di, Mehmet + te ıstakoz + u
Ahmet fish + acc. cook + past Mehmet + and lobster + acc.
'Ahmet cooked the fish and Mehmet (cooked) the lobster'

Ahmet balığ + ı, Mehmet + te ıstakoz + u pişir + di
'Ahmet (cooked) the fish, and Mehmet cooked the lobster'

Most SOV languages (e.g. Japanese) do not allow 'forward gapped' structures like the first one above. Interestingly enough, Turkish itself does not allow such structures when they are embedded:

(Ben) [Ahmed + in balığ + ı, Mehmed + in + de ıstakoz + u
pişir + diğ + in] + i bil + iyor + um
'I know that Ahmet (cooked) the fish and Mehmet cooked the lobster'

*(Ben) [Ahmed + in balığ + ı pişir + diğ + in + i, Mehmed + in + de
ıstakoz + u] bil + iyor + um
'I know that Ahmet cooked the fish and Mehmet (cooked) the lobster'

This concludes our overview of the syntax of Turkish.

Bibliography

For the classification of the Turkic languages, reference may be made to the contributions in Deny et al. (1959), to Johanson (1998), and to Schönig (1997). The following works are useful for the historical background to Turkish: Karamanlıoğlu (1972) – an overview of some historical literature, offering the author's own views on the development and geographical typology of the Turkic languages, especially those closely related to Turkish, and a discussion of the language reform – Von Gabain (1963) and Mansuroğlu (1954). Lewis (1999) is a perceptive, interesting and somewhat controversial account of the language reform.

Lewis (2000) is a revised and enlarged edition of the original grammar published in 1967. It is a comprehensive and detailed treatment of Turkish grammar with useful quotations from contemporary literature and the press. Göksel and Kerslake (2005) is in a similar vein, but more extensive, with a somewhat more pedagogical focus, and with attention to intonation. Underhill (1976) is a semi-pedagogical grammar, written in an informal early generative framework. Kornfilt (1997) is a reference grammar published in a descriptive grammars series, under the general editorship of Bernard Comrie, that uses a questionnaire common to the entire series and is thus intended primarily for use by typological and theoretical linguists.

Turkish phonology (and not only vowel harmony) has proved of continual interest to generative phonologists, starting with Lees' (1961) pioneering treatment, and new solutions to various problems continue to appear regularly in the generative phonological literature.

The fullest account of the pragmatic functions of Turkish word order is Erguvanlı (1984).

Useful overview articles on basic aspects of Turkish, i.e. on its syntax, semantics, pragmatics and phonology, can be found in Boeschoten and Verhoeven (1991). Of those, Kornfilt (1991) offers discussion of generative studies on syntactic aspects of Turkish, in addition to a description of Turkish syntax. Underhill (1986) consists of brief capsules on studies in English on numerous aspects of Turkish. Johanson and Csató (1998) consists of useful overview articles on Turkic as a language family as well as on individual Turkic languages.

References

Boeschoten, H. and Verhoeven, L. (eds) 1991. *Turkish Linguistics Today* (E.J. Brill, Leiden)

Deny, J., Grønbech, K., Scheel, H. and Togan, Z.V. (eds) 1959. *Philologiae Turcicae Fundamenta*, vol. 1 (Steiner, Munich)

Erguvanlı, E.E. 1984. *The Function of Word Order in Turkish Grammar* (University of California Press, Berkeley, Los Angeles and London)

Göksel, A. and Kerslake, C. 2005. *Turkish: A Comprehensive Grammar* (Routledge, London and New York)

Johanson, L. 1998. 'The History of Turkic', in L. Johanson and É.Á. Csató (eds) *The Turkic Languages* (Routledge, London and New York), pp. 81–125

Johanson, L. and Csató, É.Á. (eds) 1998. *The Turkic Languages* (Routledge, London and New York)

Karamanlıoğlu, A. 1972. *Türk Dili – Nereden Geliyor, Nereye Gidiyor* (= Hareket Yayınları, no. 46, Istanbul)

Kornfilt, J. 1991. 'Some Current Issues in Turkish Syntax', in H. Boeschoten and L. Verhoeven (eds), *Turkish Linguistics Today* (E.J. Brill, Leiden), pp. 60–92.

—— 1997. *Turkish* (Routledge, London and New York)
Lees, R.B. 1961. *The Phonology of Modern Standard Turkish* (Indiana University Press, Bloomington)
Lewis, G.L. 1999. *The Turkish Language Reform: A Catastrophic Success* (Oxford University Press, Oxford)
—— 2000. *Turkish Grammar*, 2nd revised edn (Oxford University Press, Oxford)
Mansuroğlu, M. 1954. 'The Rise and Development of Written Turkish in Anatolia', *Oriens*, vol. 7, pp. 250–64
Menges, K.H. 1959. 'Classification of the Turkic Languages, II', in J. Deny, K. Grønbech, H. Scheel and Z.V. Togan (eds) *Philologiae Turcicae Fundamenta*, vol. 1 (Steiner, Munich), pp. 5–10
Schönig, C. 1997. 'A New Attempt to Classify the Turkic Languages', *Turkic Languages* vol. I, pp. 117–33.
Slobin, D.I. and Zimmer, K. (eds) 1986. *Studies in Turkish Linguistics* (John Benjamins, Amsterdam and Philadelphia
Underhill, R. 1976. *Turkish Grammar* (MIT Press, Cambridge, MA)
—— 1986. 'Bibliography of Modern Linguistic Work on Turkish', in D.I. Slobin and K. Zimmer (eds), *Studies in Turkish Linguistics* (John Benjamins, Amsterdam and Philadelphia), pp. 23–51
Von Gabain, A. 1963. 'Die Südwest-Dialekte des Türkischen', in *Handbuch der Orientalistik*, I Abt., 5 Band, 1 Abschn (E.J. Brill, Leiden), pp. 174–80