

# A guide for expatriates learning to read Dari

Second Edition

Adam Baker

September 5, 2015

The purpose of this booklet is to serve as an introduction to the written form of the Dari language, for people who have already acquired some competence in spoken Dari. Learning to read Dari is helpful for a variety of reasons, not least in that it helps one to master new vocabulary that might not be reinforced in daily conversations.

The best way for an expatriate begin to read Dari is with the excellent *Let's Become Literate* (بیائید خواننده «شویم») primer and the accompanying story books (available for purchase from LCP). A person with patience and strong intuition can learn to read from that book directly, either with the help of a language teacher or a literate Afghan friend who has a little knowledge of grammar. A more structured approach is available in the LCP publication, *A workbook for reading and writing Dari*, which introduces the conventions of the alphabet, along with writing and listening exercises. The present resource can be used as a supplement to either of those approaches. It provides a big-picture overview of the Dari alphabet, with explicit instruction on certain important points. In learning to read Dari, it's possible to become lost in the details, but having an overview of the alphabet and writing conventions can prevent that from happening.

As with learning the spoken language, there is absolutely no substitute for time spent on task: only by reading can one become a good reader. Learning the facts *about* reading Dari that are presented here is no substitute for actually reading Dari, any more than reading a book about riding a bicycle is a substitute for actually learning to ride.

## 1 Why should a foreigner learn to read Dari?

Learning to read Dari is not a small undertaking, so it is appropriate here to consider a few of the many motivations that one might have for learning to read Dari.

**Learning to read Dari introduces more formal speech styles.** Some people think that the Dari spoken on the street is the “true” form of Dari, and that the literary language is something else entirely. In fact, things are not so cleanly cut. The written form of Dari is the model for the spoken speech style. This is the variety of Dari that is encountered in situations involving any formality at all: formal speeches, the way people talk on the radio or television, when people are reading aloud, and even customer service messages on a cell phone. For instance, if with a certain phone company one calls a phone that's turned off, the following message is played: [ʃumare ke ʃumal dajl namuda ed xamuʃ mebaʃad, lutfan badan tamas bigired]. This is hardly recognizable as Dari to someone acquainted only with the spoken form: the word choice and the inflections are based on the written forms. Given familiarity with the formal (written) forms, the meaning of the message becomes evident: “The number you have dialed is silent. Please get in touch later.” Thus, familiarity with the written forms increases one's ability to interact in the spoken language as well.

**Learning to read Dari helps in learning abstract vocabulary.** Being able to read in Dari opens up new opportunities for language study. It is difficult to learn abstract terms well, because they're not used very often in daily conversation. Reading printed texts can be helpful in mastering abstract vocabulary, since one can rehearse it as often as one wishes in reading.

**Learning to read Dari helps in learning relevant technical vocabulary.** A language teacher can't teach vocabulary that s/he isn't familiar with. To learn more technical vocabulary—say, related to professional activities—it can be helpful to read printed texts. As with the previous point, one can rehearse newly learned vocabulary much more easily in print than with a language teacher.

**People who can read can use a dictionary.** There are many good bilingual Persian dictionaries available. All of these, of course, are oriented toward the written form of Persian. Learning to read makes these resources accessible.

**The differences between spoken and written Dari are concentrated in common forms.** Once one learns the basic differences, moving between the spoken and written forms is not difficult. It's *not* an ongoing effort to learn a new language.

**Learning to read in a new language is easier.** A maxim of literacy is that a person only learns to read once. That is, once one has learned the idea that a letter can correspond to a sound, one has done much of the hard work of learning to read. Learning to read a second language is much easier. Dari has a reasonably clear writing system. It is slightly more complicated than Spanish, for instance, but certainly much less opaque than English. The experience of the author is that learning to read a cursive, right-to-left script is not as difficult as he thought it would be.

**Learning to read Dari makes it easier to learn to read other languages, like Pashtu.** The greater part of the difficulty in learning to read Dari is learning to read the cursive Arabic alphabet. Since this same script is used for other languages—and particularly, for all other languages in Afghanistan—learning to read in other languages is easier. Pashtu has only a few letters that are not in Dari.

**Persian has thousands of years of literary history.**<sup>1</sup> In spite of currently high rates of illiteracy in Afghanistan, Dari is a language of literature. To be illiterate in Dari is to miss a significant portion of the language, and therefore to be ignorant of a significant portion of the culture.

With these motivations in mind, the new reader should be ready to plunge enthusiastically into the study of written Dari. The nature of the task is that most of the difficulty lies in the early lessons. Therefore, perseverance is required at the beginning. After a certain point, however, the only difficulty is the occasional unfamiliar word.

## 2 The Alphabet

Dari is written with a modified version of the Arabic alphabet, which presents three challenges to new learners.<sup>2</sup> First, the Arabic alphabet is written from right to left, instead of from left to right. Second, the letters are often joined together; it is a matter of practice to be able to pull the letters apart in one's head. Third, some vowel are not written. This turns out not to be as much of a problem as one might think it would be, though of course there is a learning curve.

There are thirty-three Dari letters to learn. They are shown in Table 1. A few observations:

- A Dari letter is composed of “base shape” plus dots above or below. Several letters have the same base shape and are differentiated only by their dots (e.g., «ب», «پ», «ت», and «ث»). One, two, or three dots are used; these are placed either above or below the base shape. The dots are part of the letter, so it is important to note their number and location.
- Each letter has a characteristic sound. With only four exceptions, each letter always makes the same sound.
- Each letter has a mnemonic name, just as in English or any other language. This how people refer to the letters aloud, for instance when spelling a word. For instance «ب», which makes the [b] sound, is called [bɛ].
- Each letter is shown with a ‘half’ and ‘full’ form (in Dari, «نیم» [nim] and «پر» [pur], respectively). These are the forms that a letter takes when it is written in various parts of the word. The half forms occur at the beginning and middle of a word. The full forms occur elsewhere. To be more complete, the shape of a letter can be said to have an isolated, initial, medial, and final form. This is the nomenclature used below, though it is worth noting that the final form is usually just the full form with a connecting line, and that the medial form is usually just the initial form with a connecting line. The different shapes of a letter are called “contextual variants.”
- Generally, the contextual variants resemble one another. There are only a handful of letters for which the variants are particularly different.
- The Dari alphabet has an order, which is used for alphabetization. Some tips for learning the order, which should be done eventually, are given in Appendix A.

<sup>1</sup>“Persian” here refers to all of the spoken varieties of modern Persian (Farsi, Dari, Tajiki), and by extension to earlier forms of the language, which are preserved in written records.

<sup>2</sup>People who have previously learned to read Arabic need to make some adjustments. Dari does not have contrastive vowel length, so there is no correspondence between the length of the vowel and whether it is written or unwritten; instead, some vowels are simply not written. Many consonants that have different pronunciations in Arabic have a single pronunciation in Dari. There are also some new symbols to learn, since Dari has certain consonant sounds that Arabic does not.

Dari Name	IPA Name	Glassman	IPA	‘Half’	‘Full’
الف	alɛf	A or a	ʌ or a	ا	ا
ب	bɛ	b	b	ب	ب
پ	pɛ	p	p	پ	پ
ت	tɛ	t	t	ت	ت
ث	sɛ	s	s	ث	ث
جیم	dʒim	j	dʒ	ج	ج
چ	tʃɛ	ch	tʃ	چ	چ
ح	h	H	h	ح	ح
خ	xɛ	kh	x	خ	خ
دال	dʌl	d	d	د	د
ذال	zʌl	z	z	ذ	ذ
ر	rɛ	r	r	ر	ر
ز	zɛ	z	z	ز	ز
ژ	ʒɛ	zh	ʒ	ژ	ژ
سین	sin	s	s	س	س
شین	ʃin	sh	ʃ	ش	ش
صاد	sʌd	s	s	ص	ص
ضاد	zʌd	z	z	ض	ض
طوی	tuj	t	t	ط	ط
ظوی	zuj	z	z	ظ	ظ
عین	ajn	—	—	ع	ع
غین	ɣajn	gh	ɣ	غ	غ
ف	fɛ	f	f	ف	ف
قاف	qʌf	q	q	ق	ق
کاف	kʌf	k	k	ک	ک، ک
گاف	gʌf	g	g	گ	گ
لام	lʌm	l	l	ل	ل
میم	mim	m	m	م	م
نون	nun	n	n	ن	ن
واو	wʌw	U or w	u or w	و	و
ه	hɛ	H or e	h or e	ه	ه
ی	jʌ	I or y	i or j	ی	ی
ذ	hamza	—or ‘	—or ?	ذ	ذ

Table 1: The entire Dari alphabet, in alphabetical order, with the various forms of the letters, and pronunciations in IPA and Glassman script. Start reading on the right.

## 2.1 Connecting and Non-connecting letters

The Dari alphabet is divided into connecting letters and non-connecting letters. Connecting letters always join to a following letter in the word. When it connects to a following letter, its shape changes. The connected form of the letter—which is always narrower—is called the [nim] ‘half’ form; the unconnected form is called the [pur] ‘full’ form. On the other hand, non-connecting letters *never* join to a following letter. «ن» [nun] is an example of a connecting letter: it always connects to a following letter in the same word. «ا» [alef] is an example of a non-connecting letter: it never connects to a following letter. The example below shows how this works out for the word «نان» [nan] ‘bread’, which is spelled ن-ا-ن [nun-alef-nun].

(1) ن + ا + ن = نان

«ن» [nun] is a connecting letter, and since there is an «ا» [alef] following it, the «ن» [nun] connects to the «ا» [alef], changing its shape in the process: نا. But «ا» [alef] is a non-connecting letter, so it does not connect to the second «ن» [nun]. The second «ن» [nun] does not have any letter after it, so it is written in its isolated form. It should be kept in mind that a connecting letter always connects to the following letter, even if that letter is itself non-connecting. In the example above, the first «ن» [nun] connects to the «ا» [alef], even though «ا» [alef] is a non-connecting letter. “Non-connecting” means only that a letter does not connect to the next letter following it.

The majority of letters in Dari are connecting. The non-connecting letters of Dari are:

(2) ا د ذ ر ز ژ و

A connected script such as Dari’s poses challenges for people who are used to reading Latin script. The first challenge is that there are more forms of the letters to recognize. In Latin script, letters have two shapes: the upper and lower cases. There are more potential shapes of Dari letters.<sup>3</sup> For the most part, however, these letter shapes resemble one another very closely.

A second difficulty is that the *widths* of letters varies in Dari. A ‘t’ in Latin script is the same size whether it occurs at the beginning, middle, or end of a word (e.g., ‘tack’, ‘atom’, and ‘cat’). But in Dari, the widths of the letters varies considerably: the initial «ب» [be] in «بد» [bad] ‘bad’ is much narrower than the final «ب» [be] in «سب» [seb] ‘apple.’ For this reason, although it is tempting to think of the isolated form of the Dari word as “basic,” it is actually better to think of the initial form of the Dari letter as the basic one. If one is trained to recognize «س» as the letter [sin], then it is more difficult to recognize it when it occurs word-initially—as in the word «سگ» [sag] ‘dog’. That is because one of the most visually prominent features of the isolated form «س» [sin] is the “swash” at the end of the letter. But in the initial and medial positions there are no swashes. If «س» [sin] is learned as the basic form, however, it is still easy to recognize it at the end of a word, for instance in «کس» [kas] ‘someone’. The initial and medial forms are also much more frequently encountered in texts. It takes some time for the eye to get used to recognizing the letter shapes with their different widths, however.<sup>4</sup>

Finally, it should be emphasized that in Dari, not every break in the cursive script corresponds to a word break.<sup>5</sup> Single words such as «نان» or «کمر بند» have “gaps” in them, but are still just one word. Spaces inside of words are narrower than spaces between words. It is natural for a beginner to have difficulty in identifying where words begin and end.<sup>6</sup>

## 2.2 Connecting consonants

The connecting consonants are described below in groups, according to the basic shapes of the letters. The letters «ب» [be], «پ» [pe], «ت» [te], and «ث» [se] are identical except for the position and number of their dots. The contextual variations in the base shapes of the letters for these four letters are all the same: if one can recognize a «ب» [be] in the middle of a word, one can also recognize a «پ» [pe].

<sup>3</sup>Note that the two languages are not quite parallel. In Latin-script languages, upper case letters mark the beginnings of sentences and certain kinds of words. In Dari, the initial/medial/final/isolated forms are used irrespective of what kind of word is being written: only the position of the letter in the word is significant.

<sup>4</sup>A tangential note on word length: Latin-based ideas of what a short and long word look like do not carry forward in reading Dari. This is because of varying widths of letters, and also because some vowels are not written. The words «سب» and «حرکت» have a similar width on the printed page, but the former has one syllable ([seb] ‘apple’) and the latter has three ([harakat] ‘movement’).

<sup>5</sup>This is unlike cursive scripts in the Latin alphabet: for those, all letters are connecting, so every gap corresponds to a word break.

<sup>6</sup>Unfortunately, in printed text this difficulty is often exacerbated, with spaces inside of words that are as wide as the spaces between words. It is sometimes assumed that this is a problem with the Arabic script, but in fact it is merely poor font design. The problem never arises in a handwritten text.

This is a good time to start learning to recognize and name the letters. It is important to pay attention to the dots. One wouldn't want to learn «ب» [be] as “the letter that's shaped like a bowl,” because that's an equally good descriptor of «پ» [pe], «ت» [te], and «ث» [se] as well.

### 2.2.1 The ب-group

The ب-group is composed of the letters «ب» [be], «پ» [pe], «ت» [te], and «ث» [se]. These correspond respectively to the sounds [p], [b], [t], and [s]. Of these, «ث» [se] is encountered only very infrequently.

The various forms of «ب» [be] (representing of the ب-group) are shown below, written in isolation and also in context. At this point the other letters are unfamiliar, but try to identify the «ب» [be] in each word. The pronunciation is given for each word, but it's not important to pay attention to that; it just bothers some people not to have it.

	Final	Medial	Initial	Isolated
(3)	ب اسب	ب تبر	ب بند	ب حساب
	[asb]	[tabar]	[band]	[hesAb]

The isolated variant is written when the letter follows a non-connecting letter (in the «نان» [nAn] example above, this is the «ا» [alef]).<sup>7</sup> The initial form occurs when the letter is at the beginning of a word, or if the letter follows a non-connecting letter. For instance, in the word «کمر بند» [kamarband], an initial «ب» [be] is used because the preceding letter («ر») is non-connecting. The medial form occurs when the letter follows a connecting letter, and is followed by a letter. The final form occurs at the end of a word, following a connecting letter.

This group also illustrates the principle that it is better to think of the initial form as basic, with the isolated (or final) forms having an extraneous swash to complete the letter. Learning «ب» [be] as the basic form is not helpful, since the letter has such a different appearance (i.e., is much narrower) in the initial and medial forms. The single vertical stroke in the initial and medial forms is called a [dandan] ‘tooth.’

### 2.2.2 The ج-group

The ج-group consists of the letters «ج» [dʒim], «چ» [tʃe], «ح» [he], and «خ» [xe], which correspond to the sounds [dʒ], [tʃ], [h], and [x]. The positions of letters in the ج-group are shown below.

	Final	Medial	Initial	Isolated
(4)	ج پنج	ج لهجه	ج جیم	ج تاج
	[pandʒ]	[lahdʒa]	[dʒim]	[tAdʒ]

Here again, it is advantageous to view the initial/medial forms as basic, and the “tail” as something required just for the isolated and final forms. Try not to focus on the tail of this letter, since it doesn't occur in all positions.

The [h]-sound is commonly dropped in spoken Dari, so the presence of «ح» [he] in the written form can be surprising. For instance, in the word «امتحان» [emteħAn] ‘test’, the spoken realization of the «ح» [he] is more like a [j]: [emtejAn].

### 2.2.3 The س-group

The two letters of the س-group are «س» [sin] and «ش» [ʃin], which represent the sounds [s] and [ʃ]. Previously it was seen that «ث» [se] also represents the sound [s]; «س» [sin] is much more commonly used, however.

	Final	Medial	Initial	Isolated
(5)	س پس	س عسل	س سیر	س داس
	[pas]	[asal]	[sir]	[dAs]

<sup>7</sup>The isolated form also appears when the letter is being referred to, such as “the letter «ب».”

The challenge in identifying these letters is that the “teeth” in the letters also resemble the medial forms of the ب-group letters. Identifying individual letters in words such as «سبب» [seb] ‘apple’ (س + ب + ب) and «بيبينم» [bibinem] ‘we see (subj.)’ (ب + ي + ب + ي + ن + ي + م) is challenging at first.

#### 2.2.4 The ص-group

The letters «ص» [sʌd] and «ض» [zʌd] are fairly rarely encountered. «ص» [sʌd] is an uncommon representation of [s]—for which «س» [sin] is more commonly seen. «ض» [zʌd] represents [z], for which «ز» [ze] is more commonly used. These letters are used in certain words of Arabic origin.

	Final	Medial	Initial	Isolated
(6)	ص مَشْخَص	ص اِقْتَصَاد	ص صَابِر	ص خُصُوص
	[moʃaxas]	[ɛqtɛsʌd]	[sʌbɛr]	[xusus]

#### 2.2.5 The ط-group

The letters «ط» [tuj] and «ظ» [zuj] represent [t] and [z], respectively. As with the previous group, these are the less common representations of these sounds, occurring mainly in words of Arabic origin. The letters are virtually identical in all positions.

	Final	Medial	Initial	Isolated
(7)	ط تَوْسُط	ط مَنْطَقَة	ط طَالِب	ط مَرْبُوط
	[tawasut]	[mantɛqa]	[tʌlɛb]	[marbut]

#### 2.2.6 The ع-group

The letters «ع» [ajn] and «غ» [ɣajn] form the ع-group. «ع» [ajn] is commonly encountered, but it is often silent. In careful speech, it may be pronounced as a glottal stop ([ʔ]; the middle sound in *uh-oh*). The letter «غ» [ɣajn] is the [ɣ] sound. As can be seen below, the contextual variants of this letter are fairly distinct.

	Final	Medial	Initial	Isolated
(8)	ع جَمْع	ع بَعْد	ع دَعَا	ع دَفَاع
	[dʒam]	[bad]	[duwʌ]	[dɛfʌ]

#### 2.2.7 The ف-group

«ف» [fɛ] and «ق» [qʌf] correspond to the sounds [f] and [q].<sup>8</sup> While these letters are grouped here, it is notable that the isolated and final forms of the letters differ, in that «ق» [qʌf] has a rounder and shorter “tail,” while that of «ف» [fɛ] is longer and less curved.

	Final	Medial	Initial	Isolated
(9)	ف ضَعِيف	ف سَفِيد	ف فَكْر	ف بَرْف
	[zajif]	[safed]	[fɛkr]	[barf]

<sup>8</sup>A silly mnemonic to remember which is which: [f] is made with the lips; [q] is made in the back of the throat. There is only one pair of lips (one dot in «ف» [fɛ]), but two tonsils in the back of the throat (two dots in «ق» [qʌf]).

### 2.2.8 The ک-group

«ک» [kʌf] and «گ» [gʌf] represent the sounds [k] and [g].

	Final	Medial	Initial	Isolated
(10)	ک یک	ک فکر	ک کابل	ک سرک
	[jak]	[fekr]	[kʌbul]	[sarak]

For «ک» [kʌf] (but not for «گ» [gʌf]), there are alternate isolated and final forms. These are occasionally seen, though the forms above are more common.

	Final	Isolated
(11)	ک یک	ک سرک
	[jak]	[sarak]

### 2.2.9 Other connecting consonants

The remaining connecting consonants all have unique base shapes. They are «ل» [lʌm], «م» [mim], «ن» [nun], «ه» [he], and «ء» [hamza].

«ل» [lʌm] makes the [l] sound. In reading this letter, it is important not to confuse it with «ا» [alef], which is similar in that it has a single tall vertical stroke. The key to differentiating the two letters is that «ل» [lʌm] is a connecting letter, so in the initial and medial positions it *always* connects to a following letter, whereas «ا» [alef] *never* does. In the final and isolated forms, «ل» [lʌm] has a tail, whereas «ا» [alef] does not.

	Final	Medial	Initial	Isolated
(12)	ل کابل	ل کلید	ل لمس	ل بال
	[kʌbul]	[kelid]	[lams]	[bal]

One final complication for «ل» [lʌm] and «ا» [alef] is that when the two letters connect, they are written with a special ligature, a single letter that represents both the original ones. The lam-alef sequence is written «لا» in isolation, or «لا» when it is joined. For instance, the name of the letter «ل» [lʌm] is written «لام». The word [halʌl] is written as «حلال».<sup>9</sup>

«م» [mim] makes the [m] sound; its variants are shown below.

	Final	Medial	Initial	Isolated
(13)	م کلم	م ایمان	م من	م گرم
	[kalam]	[imʌn]	[man]	[garm]

«ن» [nun] makes the [n] sound; its variants are shown below. As seen below, in the initial and medial positions this letter acts like it belongs in the ب-group, but the medial and final forms are quite different.

	Final	Medial	Initial	Isolated
(14)	ن تن	ن بینی	ن نو	ن دان
	[tan]	[bini]	[naw]	[dan]

<sup>9</sup>The lam-alef ligature is not used in the *Let's Become Literate* books, apparently for the benefit of new readers. The author is not aware of any other publication in which the ligature is not used.

The letter «ه» [hɛ] alternately represents a vowel or a consonant. In word-initial or word-medial position, it represents the sound [h]. At the end of the word, «ه» [hɛ] can either be [h] or [a]. The former is illustrated by «کوه» [koh] ‘mountain’; the latter is illustrated by «خانه» [xɒnɒ] ‘house.’

	Final	Medial	Initial	Isolated
	ه	ه	ه	ه
(15)	مجله	بوتها	هارون	فروشگاه
	[madʒɒlɒ]	[buthɒ]	[hɒrun]	[foruʃgɒh]

There is an alternate medial form of «ه» [hɛ] that is often used in handwriting, and occasionally seen in print: instead of «بوتها» [buthɒ], for instance, «ه» [hɛ] is written as «بوتها».

«ه» [hamza] is an infrequently encountered sound that is not pronounced in casual speech; in formal speech it is pronounced as a glottal stop ([ʔ]; the middle sound in *uh-oh*). The sound appears either as a diacritic above a «پ»-like initial or medial form, or is placed directly over another letter.

	As a diacritic	Medial	Initial
	ء	ء	ء
(16)	مؤلف	مسئول	بیائید
	[muɒɒɒ]	[masul]	[bijɒjed]

## 2.3 Non-connecting consonants

### 2.3.1 The د-group

The sounds «د» [dɒ] and «ذ» [zɒ] make the [d] and [z] sounds. «د» [dɒ] is the only way to write [d], but «ذ» [zɒ] is a less common way to write [z]. The medial/final form has a somewhat different shape than the initial/isolated form, as shown below.

	Final/Medial	Initial/Isolated
	د	د
(17)	کلید	دان
	[kelid]	[dɒn]

### 2.3.2 The ر-group

The ر-group is comprised of «ر» [rɛ], «ز» [zɛ] and «ژ» [ʒɛ]. «ر» [rɛ] makes the [r] sound. «ز» [zɛ] makes the [z] sound; it is the most common letter used to represent this sound. «ژ» [ʒɛ] makes the [ʒ] sound. For these letters the medial/final and initial/isolated forms are nearly identical.

	Final/Medial	Initial/Isolated
	ر	ر
(18)	سرک	رد
	[sarɒk]	[rɒd]

This concludes the discussion of the consonant letters. Since the vowels are a little more complicated, it would be worth taking a break at this point to review the consonants: their shapes, their names, and their sounds. It is a good idea have a working knowledge of the consonants before moving on to the next section.



### 3 Vowels

While the correspondences between the consonant letters and their sounds is straightforward, vowels are somewhat more complicated. The Arabic script provides three vowel symbols: «ا» [alef], «و» [waw], and «ی» [ja]. In Dari, there are eight simple (monophthongal) vowels and five diphthongs. As a result, many vowels are not written. Approximately: vowels other than [ʌ], [i], and [u] are not written. The reader must infer when a vowel is present. When an experienced reader sees the word «سرک», which has no vowel letters, s/he knows intuitively that the word is [sarak] ‘street’. For a beginner, this may seem like a recipe for chaos, since «سرک» might also be pronounced as [sorok], [sark], [sɛrek], [sɛrk], etc.; in practice, it is not that bad, because there’s usually only one actual word that the letters can spell.

First, we review the positional variants of the three vowel letters. Of the three, «ا» [alef] and «و» [waw] are non-connecting, and «ی» [ja] is connecting.

	Final/Medial	Initial/Isolated
(19)	ا با [bʌ]	ا الف [alef]

	Final/Medial	Initial/Isolated
(20)	و نو [tu]	و والی [wali]

The shapes of «ا» [alef] and «و» [waw] are nearly identical in either context. Recall from the discussion of «ل» [lam] that «ل» [lam] and «ا» [alef] look similar, but can be differentiated in two ways: in the initial/medial forms, «ل» [lam] connects to a following letter but «ا» [alef] does not; in final/isolated forms, «ل» [lam] has a tail, but «ا» [alef] does not.

	Final	Medial	Initial	Isolated
(21)	ی بینی [bini]	ی سیب [seb]	ی یک [jak]	ی آشپزی [ʌʃpazi]

«ی» [ja] looks like a ب-group letter in the initial and medial forms. It has a more exciting variant for the final and isolated contexts. Note that in the final and isolated forms, there are no dots underneath the «ی» [ja].<sup>10</sup>

Turning now to the pronunciation of these letters, it is important to note that all three vowels letters have multiple possible sounds. This is different to the consonants, which are nearly all consistent. Interpreting the vowel letters is one of the more difficult parts of learning to read Dari.

#### 3.1 The vowel «ی»

«ی» [ja] can be the consonant [j], as in «یک» [jek] ‘one,’ the vowel [i], as in «کلید» [kelid] ‘key.’ How can these distinct pronunciations be distinguished? In the *middle* of the word, there is no way to tell, unless you know the word. «کلید» could be [kelajad], but it happens that it’s [kelid].

Things are different at the beginning of a word, however. When the sound [i] occurs at the beginning of a word, it is represented by «ای». The «ا» [alef] shows that the word begins with a vowel sound. Thus [iman] ‘faith’ is written «ایمان». If there is no «ا» [alef] at the beginning, the letter is making the consonant sound [j]. As noted above, [jek] ‘one’ is written «یک».

As an additional complication, «ی» [ja] is also used to represent the vowels [e] and [aj]. For instance «ی» [ja] is used to write «سیب» [seb] ‘apple.’ This sound is relatively infrequent in Dari. The sound [aj] is more common, as in the word «حیوان» [hajwan] ‘animal.’<sup>11</sup>

<sup>10</sup>This is a difference from the Arabic script, which has the dots in all contexts. For instance, «آشپزی» [ʌʃpazi] would be written instead of «آشپزی». The latter is the preferred Persian, but the former is often seen because of careless typesetting or software design.

<sup>11</sup>This makes sense, since the [a] is unwritten, and the [j] is written with «ی» [ja].

To summarize, «ی» [jɒ] can be either [j], [i], [e], or [aj]. If it occurs at the beginning of a word, it is always [j]; a vowel sound at the beginning of a word is written «ای». In the middle of the word, you just have to know the pronunciation of the word. The vowel can be either [i], [e], or [aj], and again it is just a matter of knowing the correct pronunciation of the word, based on your knowledge of the spoken language, or with the help of a dictionary.

### 3.2 The vowel «و»

The situation with «و» [waw] is closely parallel to that of «ی» [jɒ]. It can be the consonant [w], or one of the three vowels [u], [o], or [aw].

At the beginning of a word, «و» [waw] is always [w], as in «واحد» [wɒhed] ‘Wahed.’ If a word begins with one of the vowels, it is written «او», as in «او» [u] ‘him/her.’

«و» [waw] can also be used to write [o], as in «چوب» [tʃob] ‘wood.’ This vowel doesn’t come up very frequently, though. «و» [waw] can also be used to write [aw], as in «سودا» [sawdɒ] ‘goods.’<sup>12</sup>

In parallel with «ی» [jɒ], then, «و» [waw] can be used to write either a consonant or a vowel. If «و» [waw] occurs at the beginning of a word, it is a consonant; at the beginning of a word the vowel is written «او». If it occurs elsewhere, it is a matter of recognizing the word. As a vowel, «و» [waw] can be either [u], [o], or [aw].

### 3.3 The vowel «ا»

The vowel «ا» [alef] in the middle of the word makes the [ʌ] sound, as in «نان» [nɒn]. At the beginning of the word, things are quite different. As discussed previously, «ا» [alef] can be used to write [i] and [u] at the beginning of a word. When «ا» [alef] is written by itself at the beginning of a word, it can be either [a], [ɛ], or [ʊ]. For instance [ɛsm] ‘name’ is written «اسم», [abr] ‘cloud’ is written «ابر», and [ʊstɒd] ‘master’ is written «استاد». One can tell that these are representing one of these three sounds—though not *which* one—because the following letter is not «ی» [jɒ] or «و» [waw].

To illustrate these principles, consider the word «انار». One can’t be certain whether the first «ا» [alef] is [a], [ɛ], or [ʊ], but the second «ا» [alef] *must* represent [ʌ]. In this case, the word is [anɒr] ‘pomegranate’.

How is a word written if it begins with the [ʌ] sound? In that case the sound is written with a different letter, «آ». The word [ʌf] ‘noodle’ is written «آش». «آ» is called [alef mad].

To fill in the picture of vowels at the beginnings of words, recall from the section on consonants that some consonants are not pronounced. A word that in spoken Dari begins with a vowel might actually begin with a consonant, such as «ه» [hɛ], «ح» [hɛ], or «ع» [ajɒ]. Examples of this are «عبور» [ʊbur] ‘transition’, «عزیز» [aziz] ‘Aziz’, «هوشیار» [(h)ʊʃjɒr] ‘intelligent, savvy’, and «حساب» [(h)ɛsɒb] ‘arithmetic’.

## 4 Remaining complications

- The ezafa suffix joins the words in a noun phrase, e.g., «بوت عزیز» [but-ɛ aziz] ‘Aziz’s shoe’. This suffix is generally not written, but there are three exceptions. If a word ends in «ه» [hɛ], then the suffix is written with a diacritic «ء» [hamza]: «خانه من» [xɒne mɒn]. The second exception is that if a word ends in «ا» [alef] or «و» [waw], the suffix is written with a «ی» [jɒ]. Examples: «آلובالو» [ʌlubɒlu] ‘cherry’ and «آلوبالوی شیرین» [ʌlubɒlu-je ʃirin] ‘sweet cherry’; «خانههای من» [xɒnahɒ-je mɒn] ‘my houses.’ The final exception is that if a word ends in a «ی» [jɒ], the [ɛ] suffix is not written at all: «رحصتی» [ruxsati] ‘holiday’ and «رحصتی عمومی» [ruxsati-je umumi] ‘general holiday’.
- There are a few words in which the [aw] sound at the end of a word is written with a «ب» [bɛ]. This reflects the educated or more Farsi-like pronunciation. Examples are «آب» [aw] ‘water’ and «شب» [ʃaw] ‘night’ (on the «و» [waw] in the latter example, see below). This «ب» [bɛ] is occasionally pronounced in spoken Dari.
- In some words an initial [xɒ] sound is written «خوا». Examples are «خواب» [xɒwɒ] ‘sleep’ and «خواندن» [xɒndan] ‘to read’. In very formal speech one might hear [xwɒb] or [xwɒndan].
- In Dari the [an] suffix changes a word into an adverb (like the English -ly suffix). For instance, [mʊtɒbeq] ‘corresponding’ becomes [mʊtɒbeqan] ‘correspondingly’. There is a special spelling for this in Dari, using an «ا»-and-diacritic combination: «اَ». The previously mentioned words are spelled «مطابق» ‘corresponding’ and «مطابقاً» ‘correspondingly’.

<sup>12</sup>This is of course exactly parallel to how «ی» [jɒ] is used to write [aj].

- (Rare) At the end of words a «ی» [jɒ] sometimes makes the [ɒ] sound. An example is «معنی» [mɒnɒ] ‘meaning’.
- (Rare) Occasionally a diacritic (called [taʃdid]) is written over a consonant to indicate a lengthened pronunciation «نخار» [nadʒar]. This is not encountered consistently in print, and it does not change the pronunciation. It follows, however, that if the same letter is written twice in a row, then there must be an intervening vowel, as in «ممکن» [mumken] ‘possible’.<sup>13</sup>
- (Rare) In certain Arabic words, the letter «أ» makes the [a] sound, as in «تأكيد» [takid] ‘emphasis.’ These words are often written with a plain «ا» [alef] instead, however, giving rise to words with an «ا» [alef] in the middle of the word that makes the [a] sound: «تاكيد».

## 5 Tips for getting started

Probably the most daunting thing about learning to read Dari is coping with the lack of vowel symbols. Earlier it was noted that the word «سرک» could be read with a variety of combinations of vowels, other than the correct pronunciation, [sarak]. In practice, this ends up not being much of a difficulty. The reason is that most words are clearly differentiated by their consonants; vowels are less important. Indeed, even speakers of English commonly eliminate vowels from SMS messages, and the result is only marginally less intelligible. There are very, very few instances of Dari words that can be confused because of the absence of written vowels.

It is important to keep in mind that a mature reader does not read letter-by-letter, but word-by-word.<sup>14</sup> That is, the shapes of whole words become familiar as one reads, so that the brain is decoding the whole word, rather than taking apart each letter. This happens naturally with practice, but a new reader lacks that practice. The initial effort must be made to laboriously pick apart the letters of each word, and then to find the meaning. It’s undoubtedly tedious, but do keep in mind that it becomes automatic with practice.

With that said, new words are often encountered, and a reader needs skills to figure out what an unfamiliar word is. There is a heuristic rule that the author has found to be very useful in reading Dari: guess [a]. That is, when ones encounter a sequence of letters must have a vowel, try first to insert an [a].<sup>15</sup> For some words, like «سرک» or «حرکت», guessing [a] produces the right word ([sarak] and [harakat]). For the word «کلید», guessing [a] produces the incorrect [kalid], but still sounds a lot like [kelid].

A related problem is knowing where to put the vowels. Even non-native speakers of Dari have some intuition about what a possible Dari word is. In the example of «کلید», a reasonably fluent non-native speaker would probably not even consider the pronunciation [klid]: Dari words simply can’t begin with [kl], so there *must* be an intervening vowel. Even with these intuitions, however, it is still occasionally necessary to guess. For the word «حرکت», [harkat] is a possible Dari word—for instance, «شرکت» is pronounced [ʃerkat]—but the correct pronunciation of «حرکت» turns out to be [harakat].

The easiest words to recognize are the words known already in the spoken language. With surprisingly little practice, these words are recognized automatically. The more difficult words are the ones learnt for the first time on the printed page. Unless one reads texts from entirely within one’s sphere of previous experience, learning new words is part of learning to read. Obviously, broadening one’s vocabulary is one of the prime motivators to learn to read: only, be aware that this is also the more laborious and time-consuming part of the reading process.

It’s impossible to overstate the important of repetition and practice. A technique that the author used in learning to read was to read the same passage over and over again. While one might think that reading the same thing repeatedly wouldn’t force one’s brain to learn the symbols, in fact, the rote procedure of moving one’s eyes over the words with understanding is just what is needed to firm up reading skills. Just twelve words make up a *quarter* of the words in a typical Dari text; 190 words make up *half* of the words in a text. Being able to read these very frequent words quickly helps in reading fluently.

At the outset, it helps to adopt some practices that may seem childish: following the words with one’s finger and reading aloud. These are helpful practices, though, which is why children use them. If possible, try to find printed text that uses a larger font size than would be used for Latin script. For whatever psychological reason, an unfamiliar script looks tiny even if it is printed in a reasonable font size.

A dictionary is necessary for looking up unfamiliar words, and for checking the spellings of words one knows from the spoken language. The author strongly recommends *Farhang Moaser* («فرهنگ معاصر») Persian-English and English-Persian dictionary, compiled by Suleiman Haim («سليمان حيم»), which is widely available. Multiple editions are available; it is important to buy a copy that shows the pronunciation of the Persian

<sup>13</sup>An uncommon but systematic exception to this generalization is in compound words, where a double-letter can occur even if there is no intervening vowel. Example: «يادداشت» [jaddaʃt] ‘memo’ from the words meaning ‘memory’ and ‘to have’. (Mace 2003: 196)

<sup>14</sup>And even phrase-by-phrase!

<sup>15</sup>[a] is more than twice as common as [ɛ] and more than ten times as common as [o], so it makes more sense to guess [a] than anything else.

words. To use a dictionary, one needs to learn the order of the alphabet; see Appendix A. In looking up words, it is necessary to keep in mind that there are three sounds have multiple letters, [s], [t], and [z]. If one knows only the spoken form, there is nothing to be done but to check every possible spelling. The following recommendations for speeding up this process are based on how frequently the various letters occur. For [s], try «س» [sin], then «ص» [səd], and finally «ث» [se]. For [t], try «ت» [te], and then «ط» [tuj]. For [z], try «ز» [ze], then «ض» [zəd], then «ذ» [zəl], and finally «ظ» [zuj]. If a word begins with an unwritten vowel sound, try «ا» [aləf], then «ع» [ajn], then «ه» [he], and finally «ح» [he].

## 6 Differences between written and spoken Dari

There are not very many differences between written and spoken Dari, but the differences that are there are very frequently encountered. For a learner, this is a double-edged sword. On one hand, once the new forms are learned, then they are well-reinforced since one reads them so often. On the other hand, since the forms are frequent, at the beginning of the learning process it seems that everything is different. The key is to keep practicing reading until the new forms become familiar.

The items below are presented in list format, since each change is a fairly straightforward one. The most helpful strategy might be to skim this list to be aware of the differences; practice will reinforce each point.

### 6.1 Pronunciation differences

Pronunciation differences are not uncommon between spoken and written Dari. Remember that these are actual differences in pronunciation, and not just strange spellings. When reading Dari aloud, one's pronunciation must reflect the written form. Producing the spoken form when reading is considered a mistake.<sup>16</sup> Of course, if one's goal is merely to read Dari silently, it is not strictly necessary to learn the written pronunciations. This would preclude the possibility of being able to understand to formal speech, however. It would also be embarrassing to make so many 'mistakes' if one ever had to read a text aloud.

The list below covers the most frequent differences between the spoken and written pronunciation; others are learned naturally through experience.

- The verb [kadan] 'to do' is written «کردن» [kardan]. This affects verb forms based on the past tense stem; the present tense stem is the same as the spoken form.
- The continuous prefix [me] is sometimes written in non-joining form («می‌بینم» [mibinam] 'I am seeing'), and sometimes in joining form («میبینم»).
- The present tense stem of «رفتن» [raftan] 'to go' is «رو» [raw], as in «می‌روم» [merawam] 'I am going' (spoken [merum]).
- The present tense stem of «دادن» [dadan] 'to give' is «ده» [dah] or [deh], as in «می‌دهم» [medaham] 'I am giving' (spoken [metum]).
- The present tense stem of the word meaning 'to desire' is written, e.g., «میخواهم» [mexaham], or more formally [mexwaham] (spoken [mexahjom]). This form is predictable from previously mentioned conventions: «خو» usually represents [xʌ] at the beginning of a word, and «ه» is generally [j] in the middle of a word.
- The third person present form of 'to be' ('he/she/it is') is written with a «ت» [te] at the end: [ast] «است» (spoken [as]).
- The present tense forms of 'to be' are sometimes written with a «ه» [he] («هست» [hast]) and sometimes without («است»).
- The negative form of 'to be' is written «نیست» [nest] (spoken [nes] or [nis]).
- The preposition 'for' is written «برای» [ba'rajɛ] (spoken [bare]).
- The preposition 'in' is written «در» [dar] (spoken [da]).
- The object marker is written «را» (spoken [ra] or [a]). It is always pronounced [rʌ], no matter the context. Unlike the spoken form, the [r] is never dropped.

<sup>16</sup>Note that this is different to English, in which pronouncing an [h] in words like 'honor' or 'honesty' is a mistake.

English	Spoken	Written	Written Pronunciation
I went	raftum	رفتم	raftam
you (sg.) went	rafti	رفتی	rafti
he/she/it went	raft	رفت	raft
we went	raftem	رفتیم	raftem
you (pl.) went	raften	رفتید	rafted
they went	raftan	رفتند	raftand

Table 2: Differences in verb inflection in spoken and written Dari. The differences are illustrated here with the simple past forms, but also hold for the other tenses, aside from the present perfect, which is illustrated in Table 4.

- The plural suffix is written «ها» [hʌ] (spoken [ʌ]). Sometimes the «ه» [hɛ] is written in non-joining form «بوت‌ها», and sometimes it is written in joining form «بوتها». Some writers—unable to produce the former but aware that the latter is incorrect—write the suffix as if it were a separate word: «بوت‌ها». In pronouncing the written form, the final vowel of the stem is never lost, as it sometimes is in the spoken form: «خانه‌ها» is [xʌnahʌ] (spoken [xʌnʌ]).
- Written text usually has a greater variety of plural forms. The suffixes [ʌt] and [ʌn] are commonly seen, with the latter being used exclusively for words that refer to humans. Examples are «حيوانات» [hʌjwʌnat] ‘animals’ and «پيروان» [pʌjrawʌn] ‘followers’. More difficult to identify are ‘broken plurals’ borrowed from Arabic, which have a different shape entirely: e.g., «عمل» [amal] ‘act’ becomes «اعمال» [ɛmwʌl] ‘acts’.<sup>17</sup> Until one develops intuitions about these forms through experience, the only thing to be done is to look up the words in a dictionary.
- The demonstrative ‘this’ is written «اين» [in] (spoken [i]). The demonstrative [u] ‘that’ is written «آن» [ʌn] (spoken [u]).
- The word ‘if’ is written «اگر» [agar] (spoken [aga]). Similarly, ‘other’ is written «ديگر» [digar] (spoken [dɛga]).
- Some common function words that have an [h] in the written form are: «هر» [har] ‘every’, «هم» [ham] ‘also’, and «هيچ» [hetʃ] ‘no/none’.
- Occasionally a word with a [j] sound at the end in the spoken language does not have one in the written language. Examples are «جا» [dʒʌ] ‘place’ and «بو» [bu] ‘smell’.

## 6.2 Inflectional differences

Table 2 shows the differences in verb inflection between the spoken and written Dari forms. Spoken Dari is [raft-en] ‘you (pl.) went’, but written Dari is «رفتيد» [raft-ed], with a [d] at the end instead of an [n]. In the ‘they’ form there is also a final [d] in «رفتند» [raft-and] that is not pronounced in the spoken language. The ‘I’ forms are pronounced [raft-am] instead of [raft-um]. Since neither [u] nor [a] is written, this is not reflected in the written system, but proper pronunciation is expected when text is read (and in other formal contexts).

Another inflectional difference occurs in the present continuous and subjunctive forms. Here the ‘he/she/it’ forms take the [a] suffix in spoken Dari, for instance [mekuna] ‘he/she/it does’. The written form has a «د» [dʌ] for this suffix, pronounced [ad]: «ميكند» [mekonad].

The inflectional differences are slightly different when the verb stem ends in a vowel letter («ا» [aɛf], «و» [waw], «ي» [jʌ]). In this case a «ي» [jʌ] is added before the agreement suffix, as shown in Table 3. (As it happens, the only time a stem ends in a vowel is if it is a present tense stem.) The additional «ي» [jʌ] reflects the pronunciation of the affix in the spoken form—e.g., [ʌmad-um] ‘I came’ [mijʌ-jum] ‘I am coming’—so it is not difficult to remember the rule. When a stem ends in a vowel letter it is necessary to pay attention to whether the extra «ي» [jʌ] is there or not; otherwise the 1st person singular and plural forms (‘I’ and ‘we’) and the 3rd person singular and 2nd person plural forms (‘he/she/it’ and ‘you (pl.)’) can be easily confused. Compare «ميايد» is ‘you (pl) are coming’ with «ميايد» is ‘you (pl) are coming’.<sup>18</sup>

<sup>17</sup>In fact, the shape is not entirely different: the order of the consonant letters stays the same. Admittedly this is not much of a help, but it might help in recognizing plurals of some familiar words.

<sup>18</sup>This and other topics are covered in somewhat more detail in *A workbook for reading and writing Dari*.

English	Spoken	Written	Written Pronunciation
I am coming	mijAjum	میایم	mijAjam
you (sg.) are coming	mijAji	میایی	mijAji
he/she/it is coming	mijAja	میاید	mijAjad
we are coming	mijAjem	میاییم	mijAjem
you (pl.) are coming	mijAjen	میایید	mijAjed
they are coming	mijAjan	میایند	mijAjand

Table 3: Inflectional endings with a verb stem ending in a vowel letter (in this case «ا» [alef]).

English	Spoken	Written	Written Pronunciation
I have gone	raf'tem	رفته ام	rafta 'am
you (sg.) have gone	raf'ti	رفته ای	rafta 'i
he has gone	raft	رفته	rafta
we have gone	raf'tem	رفته ایم	rafta 'em
you (pl.) have gone	raf'ten	رفته اید	rafta 'ed
they have gone	raf'tan	رفته اند	rafta 'and

Table 4: Differences in the perfect tense in spoken and written Dari.

### 6.3 Syntactic differences

- In spoken Dari there is a somewhat uncommon way to form the future tense: by placing the helping verb [xat] before the simple past tense of the verb. For instance, [ma xat raftum] means 'I will go'. This form is more common in writing, though with changes. The verb agreement goes on the helping verb instead, and the main verb appears in the bare simple past form (the he/she/it form). In written Dari, 'I will go' is «من خواهم رفت» [man xwAham raft]. The written form of 'he will go' is «او خواهد رفت» [u xwAhad raft].
- The perfect tense is written differently; this is shown in Table 4. In the written form the perfect tense is formed with the participle (i.e., the infinitive minus [n]) plus a separate word for agreement. In the spoken these are blended together in one word. Thus spoken [raf'tem], but written «رفته ام» [rafta am].

### 6.4 Word choice differences

There are many instances of words that are used more in writing (and formal situations) than in colloquial spoken Dari. The examples below are the more frequently encountered ones.

- The verb «نمودن» [nɛmudan] is sometimes used instead of «کردن» [kardan]. All of the many compound verbs using «کردن» thus have more formal counterparts that use «نمودن».
- The verb «گشتن» [gaʃtan] is sometimes used where «شدن» [ʃudan] might be used in spoken Dari.
- Instead of [az xatɛɛ kɛ] or [bA xatɛɛ kɛ], «چون که» [tʃun ke] is more commonly seen in print.

## Acknowledgments & Bibliography

The author also gratefully acknowledges the assistance of Ustad Amir Muhammad in correcting errors and improving explanations. All remaining mistakes and inadequacies are of course the responsibility of the author.

Awde, Nicholas and Putros Samano. 2000. The Arabic Alphabet: How to Read & Write It. Lyle Stuart.  
Glassman, Eugene. 2000. Conversational Dari: An introductory language course in Dari (Afghan Persian), 5th Edition. Peshawar: Interlit Foundation.

Mace, John. 2003. Persian Grammar: For reference and revision. New York: RoutledgeCurzon.

Thackston, Wheeler M. 1993. An Introduction to Persian, Revised Third Edition. Ibex Publishers.

Latin-based	Dari	Arabic	Mnemonic
0	۰		
1	۱		looks like a 1
2	۲		has two strokes
3	۳		has three strokes
4	۴	٤	
5	۵	٥	
6	۶	٦	
7	۷		looks like the 'v' in 'seven'
8	۸		
9	۹		looks like a 9

Table 5: Persian & Arabic numerals.

## A Learning the order of the Dari alphabet

It isn't necessary to learn the order of the Dari alphabet immediately. It is useful, however, for using dictionaries or other resources whose entries are organized alphabetically. Refer to Table 1 for order of the alphabet. This appendix provides a few mnemonic devices that can be used to learn the order more simply. One may take or leave any of these items; they are merely observations that the author has found helpful.

- The letters are grouped by their base shape. First comes «ا» [alef], then the letters shaped like «ب» [be], then those shaped like «ج» [dzim], then «د» [dal], then «ر» [re], then «س» [sin], then «ص» [sad], then «ط» [tuj], then «ع» [ajn], then «ف» [fe], then «ک» [kaf]; the remaining base shapes have only one letter each.
- Within these subgroups, the first letter of the group is often the unmodified version, with the following letters being the “decorated” versions: «س» [sin] precedes «ش» [šin]. (This doesn't hold for the ب-group or the ج-group, however.)
- The «ا» [alef] – «ب» [be] order is like the Latin a-b order (and like the Greek alpha-beta order, which itself gives us the word “alphabet”). If you happen to know the term *abjad*, this can be a mnemonic for the first four letter groups: «ا» [alef] – «ب» [be] – «ج» [dzim] – «د» [dal].
- The «ک» [kaf] – «گ» [gaf] – «ل» [lam] – «م» [mim] – «ن» [nun] – «و» [waw] sequence is like the Latin k-l-m-n-o sequence. («گ» [gaf] intervenes, but it must since it is the more decorated form of «ک» [kaf].)

## B Learning the numbers

In addition to the letters of Dari, the numbers need to be learned as well. These are shown in Table 5, along with mnemonic devices for recognizing the shapes. Somewhat surprisingly, numbers in Dari are written just as they are in Latin script: from left-to-right. Thus, the number «۱۵» is 15, not 51. Note, however, that the overall order remains right-to-left. If someone were referring to a range of 15–20, it would be written «۲۰-۱۵».

The only other complication is that for the numbers 4, 5, and 6, («۴», «۵», «۶») there are alternate forms: respectively, «٤», «٥», «٦». These alternate forms are actually the Arabic varieties of these numbers. They are seen fairly commonly in Afghanistan, however.

## C The vowel accent markers

Dari can be written with accents to indicate the presence of a vowel. A slanted line under a consonant (called [zer]) indicates the presence of the vowel [ɛ] following that consonant. For the word [kelid] ‘key’:

(22) کلید

A slanted line *over* a consonant (called [zebar]) indicates the presence of the vowel [a], as in [sar] ‘head’ below:

(23) سَر

The presence of a small diacritic called [peɸ], which has the shape of a small «و» [waw], represents the vowel [u], as in [pur] ‘full’:

(24) پَر

When the diacritics are followed by vowels, this may be a diphthong, as in the words [ajnak] ‘eyeglasses’ and [sawdʌ] ‘goods’ below.

(25) عَينَك

(26) سَوَدَا

These diacritics are used only in very specific situations: in very early elementary school materials, or written text when it is necessary to indicate the pronunciation of a word (e.g., if a word might be confused with another one).<sup>19</sup> The fact that these diacritics were not introduced earlier indicates the attitude of the author toward their use. Although they seem helpful, a new reader shouldn’t depend on them. An important part of learning to read Persian is to learn to identify a word based on the sounds that are written: this can’t be learned if one is constantly relying on diacritics.

## D Ideas for early reading

After the first primers have been completed, what should a new reader begin to read? Based on the principle that in learning one moves from what is known to what is unknown, the following progression may be suggested for a new reader.

**Known subject, known text.** At the very beginning, it is most helpful to read translated texts, since these provide maximum context, thus allowing a focus on the process of decoding rather than of discovering the meaning. These may be stories that are already familiar, or that are also available in English translation. The IAM Language & Culture Programme has a number of short (1–4 page) fairy and folk tales, which are available in English and written Dari.<sup>20</sup> The organization Operation Mercy has published abridged translations of several Western classics (*Les Misérables*, *Tom Sawyer*, etc.), which are available for purchase from the Kabul office of Rahmat Publications, and in bookstores around the country. The BBC has published the well-known “BBC books,” the plots of which can largely be inferred from the pictures; these can be purchased from the BBC’s Kabul office, or from second-hand bookshops.<sup>21</sup>

**Known subject, unknown text.** The next step is to read books about familiar subjects, but in a unfamiliar text. This is a stage, for instance, in which one might wish to read books about one’s professional interests. It is not difficult to visit a bookshop, mention a topic to the proprietor, and buy a large number of reasonably priced books pertinent to one’s professional interests. Wikipedia articles written in Persian are also a good resource.<sup>22</sup> One difficulty at this stage is that, depending on the book, the language might be very elevated, and the use of a dictionary might therefore be a constant necessity.

**Unknown subject, known text.** This category covers reading material that is available in Dari and English, but which addresses subjects that the reader usually does not discuss in Dari, or at all. This could be any material which can be found in English and Dari translation. The military propaganda newspaper that usually comes with naan is printed in Dari, Pashto, and English, for instance.

**Unknown subject, unknown text.** This final category, of course, covers the rest of the printed material in the world. The most straightforward sources for new texts are newspapers, bookshops, and Afghan or Iranian media available on the internet.

<sup>19</sup>In fact, the vowel diacritics are most commonly used in the Qur’an to indicate proper pronunciation of the Arabic words. Arabic is a separate language of course, and in fact the diacritics have different interpretations in Arabic.

<sup>20</sup><http://iam-afghanistan.org/LCP>

<sup>21</sup>BBC Afghan Education Project, House 271, Street 1, Qalai Najarha, Khair Khana. A caution: the BBC books suffer from very poor typography—particularly in the spacing between non-connecting letters and words—which can make them very hard to read. Some books are also printed with a calligraphic style that is difficult for readers just learning the shapes of the letters. These are probably not appropriate for first efforts in reading.

<sup>22</sup><http://fa.wikipedia.org/>