

Language Specific Peculiarities Document for Pashto as Spoken in Pakistan and Afghanistan

1. Special handling of dialects

Pashto belongs to the Eastern Iranian branch of the Indo-European family of languages. It is principally spoken in Afghanistan and the North West Frontier Province of Pakistan. Speakers are also found in other neighboring countries, but not many Pashtuns live outside these countries.

Appen Butler Hill determined that there are four main dialect regions relevant to speech data collection, based on phoneme distributions. These broadly correspond to four geographical regions, here denoted South-West (e.g., Kandahar, the written standard), North-East (e.g., Peshawar, Jalalabad, some of Kabul), South-East (e.g., Quetta) and Central/North-West (less common variants, e.g., Central Ghilzai, Wardak, some outlying areas of Kabul). The Pashto language has considerable phonological and lexical variety, and there are significant issues relating to orthographic standardization.

Due to political instability in Afghanistan, all speakers were recruited and recorded in Pakistan for this collection. All effort was made to record speakers from the above four main dialect regions. However, it should be noted that conflicts within the region have large impacts on the migration of speakers across national borders and it is possible that some dialectal groups may be difficult to recruit within Pakistan. It was therefore not feasible to guarantee any specific dialectal distribution.

In 1936, Pashto was designated one of the two official languages of Afghanistan. The other designated language is Dari. Five other Afghan languages are also recognized. As well as Afghanistan, Pashto speakers also live in Pakistan and Iran.

Approximately 28% of the Afghan population speaks a variety of Pashto as a first language (8 million). The population of Afghanistan is estimated at 28,395,716 (July 2009).

| REGION | FIRST-LANGUAGE SPEAKER OF PASHTO | % OF POPULATION |
|-------------|----------------------------------|-----------------|
| Afghanistan | ~ 8 million | ~28% |
| Pakistan | ~ 9 million | ~12% |

Table 1: Pashto-speaking Population Estimates in Afghanistan and Pakistan

Alternative names used for this language are: Pushtu, Afghan, Pushto, Pashto, Pashtoe, Pashtu, and Pukhto. The variation in spelling of the language's name (Pashto, Pukhto, etc.) stems from different pronunciations in the various dialects; for example, the second consonant is a retroflex [s̠] in the Kandahari dialect, a velar fricative [x] in the North East dialect (Peshawari, Jalalabadi) dialect and a postalveolar fricative [ʃ] in the south-east, around Quetta. These variants alone are indicative of the general difficulties around spelling standardization.

In Afghanistan, Pashto dialects are a reflection of the underlying Pashtun tribal structure as the differences between the tribes are considered to be essentially linguistic in nature. The people are called 'Pakhtoon' in the north, 'Pashtoon' in the south. The main Pashto clans are: Mohmandi, Ghilzai, Durani, Yusufzai, Afridi, Kandahari (Qandahari), Waziri, Chinwari (Shinwari), Mangal, Wanetsi. The Ghilzai speakers are nomadic and 24% of the national population. The Durani speakers, 16%, live in permanent settlements.

Dialectal differences in Pashto are largely phonological and lexical. It is not possible to clearly divide the dialects only by tribe or by geography. The current division as described below is based on a phonology

which involves a combination of tribal affiliation and geography. The proposed dialect boundaries are approximate and often intersecting.

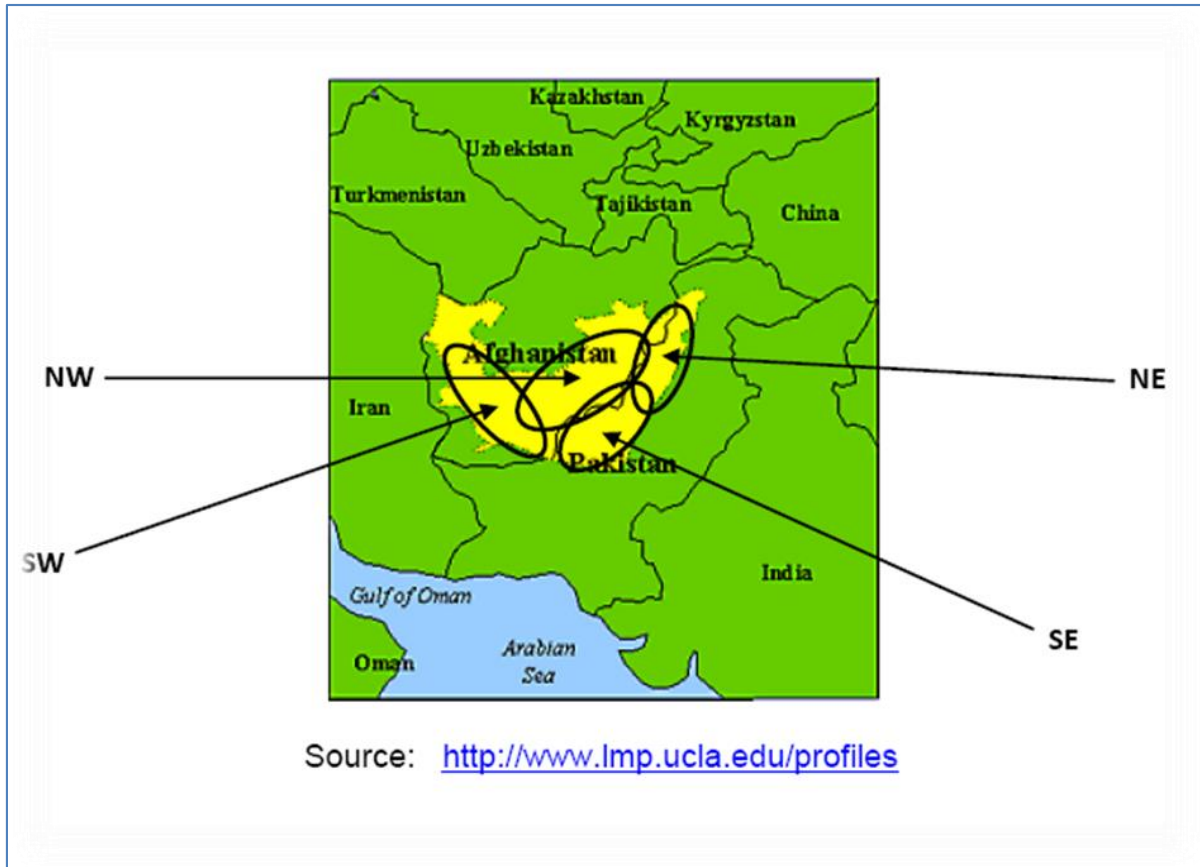


Figure 1: Distribution of the Major Dialect Divisions

1.1. Lexical variant ranking by dialect

One typical division of the dialects is into North and South varieties. In total, in all countries in which Pashto is spoken, about 9.7 million people speak Northern dialects and about 2.7 million speak Southern dialects. Most speakers of Pashto speak one of these two broad dialects. There is generally an 80% lexical similarity between the northern and southern varieties of Pashto.

The South-West (Kandahar) dialect, though less populous, is the basis of the standard written form and in this sense can be considered the standard. The South-West dialect pronunciations are therefore provided as the primary lexical variant (1), followed by North-East (Peshawar) (2), South-East (Quetta) (3), and Central/North-West which includes speakers from the Central Ghilzai and Wardak tribes (4).

A phonological division between the dialects is possible based on a distribution around 5 particular consonants. The dialectal classification proposed below is similar to and based on Josef Elfenbein, "Pashto Phonology", *Phonologies of Asia and Africa* (Indiana: Eisenbrauns, 1997). The distribution of five consonants can divide the Afghani Pashto dialects, as shown in Table 2.

| Dialect Group ¹ / Consonant (character) | (0x69a) ښ | (0x696) ږ | (0x685) څ | (0x681) ځ | (0x698) ژ |
|--|--|---|--|---|--|
| SW | <i>/ʃ/ /s̺/</i> (voiceless retroflex fricative) | <i>/ʒ/ /z̺/</i> (voiced retroflex fricative) | <i>/ts/ /tʃ/</i> (voiceless alveolar affricate) | <i>/dʒ/ /dʒ/</i> (voiced alveolar affricate) | <i>/ʒ/ /Z/</i> (voiced postalveolar fricative) |
| NE | <i>/x/ /x/</i> (voiceless velar fricative) | <i>/g/ /g/</i> (voiced velar plosive) | <i>/s/ /s/</i> (voiceless alveolar fricative) | <i>/z/ /z/</i> (voiced alveolar fricative) | <i>/dʒ/ /dʒ/</i> (voiced alveo-palatal affricate) |
| SE | <i>/ʃ/ /S/</i> (voiceless postalveolar fricative) | <i>/ʒ/ /Z/</i> (voiced postalveolar fricative) | <i>/ts/ /tʃ/</i> (voiceless alveolar affricate) | <i>/dʒ/ /dʒ/</i> (voiced alveolar affricate) | <i>/ʒ/ /Z/</i> (voiced postalveolar fricative) |
| Central/NW | <i>/ç/ /C/</i> (voiceless palatal fricative) | <i>/j/ /j/</i> (voiced palatal fricative) | <i>/s/ /s/</i> (voiceless alveolar fricative) | <i>/z/ /z/</i> (voiced alveolar fricative) | <i>/z/ /z/</i> (voiced alveolar fricative) |

Table 2: Division of Afghan Pashto Dialects by Consonant Distribution. The IPA symbols are included in blue font. The black font represents SAMPA symbols.

1.2. Geographical distribution of dialect speakers

The four dialect groups cover the following regions/tribal affiliations:

South West Group (Nearest large city = Kandahar)

Kandahar (the basis of the Afghan written standard)
Farah

Herat
South Ghilzay

North East Group (Nearest Pakistani large city = Peshawar)

Peshawar
Shinwari
northern parts of North-West Frontier Province (NWFP)
Yusufzai
Mohmand

Jalalabad
North East Ghilzay
Hazara
Bangas
Orakzay
some Afridi

South East Group (Nearest Pakistani large city = Quetta)

Southern parts of the NWFP
Sherani
Bannu
Waziri (north and south sub groups)
Baluchistan

Kakari
Acakzay
Tarin

Central/North West Group (Nearest large cities = outskirts of Kabul)

East and North East Afghanistan
Kabul outskirts
Central Ghilzay

Wardak
Ghazni

2. Deviation from native-speaker principle

No special deviation.

¹ SW = south west; NE = north east; SE = south east; Central/NW = Central/north west.

3. Special handling of spelling

The South-West (Kandahar) dialect is the most prestigious dialect. The Kandahar pronunciation is the basis of the standard written form.

This database utilizes a standard Pashto spelling system, which mirrors most closely the Kandahari dialect. Important differences in pronunciation across the dialects are captured in the SAMPA representations, rather than through spelling variation. In Pakistan, the spelling standard for Pashto is not always followed. There is a tendency to substitute the Urdu forms of letters, which fail to distinguish some of the characteristic phonological distinctions needed in Pashto, such as the five *yah* symbols. We use a standardized spelling system based on the Kandahari pronunciations for the Pashto transcription and capture the regional pronunciation differences in the Lexicon (rather than in the Transcription). Pashto characters and code points are described in Table 3.

| CODEPOINTS | Pashto | ROMAN | CODEPOINT DESCRIPTION |
|------------|--------|-------|---|
| 0x627 | ا | A | ARABIC LETTER ALEF |
| 0x622 | آ | | ARABIC LETTER ALEF WITH MADDA ABOVE |
| 0x628 | ب | b | ARABIC LETTER BEH |
| 0x67e | پ | P | ARABIC LETTER PEH |
| 0x62a | ت | t | ARABIC LETTER TEH |
| 0x67c | ټ | 8 | ARABIC LETTER TEH WITH RING |
| 0x62b | ټ | v | ARABIC LETTER THEH |
| 0x62c | ج | j | ARABIC LETTER JEEM |
| 0x681 | ح | 7 | ARABIC LETTER HAH WITH HAMZA ABOVE |
| 0x686 | چ | J | ARABIC LETTER TCHEH |
| 0x685 | ح | 6 | ARABIC LETTER HAH WITH THREE DOTS ABOVE |
| 0x62d | ح | H | ARABIC LETTER HAH |
| 0x62e | خ | x | ARABIC LETTER KHAH |
| 0x62f | د | d | ARABIC LETTER DAL |
| 0x689 | ډ | 9 | ARABIC LETTER DAL WITH RING |
| 0x630 | ذ | ^ | ARABIC LETTER THAL |
| 0x631 | ر | r | ARABIC LETTER REH |
| 0x693 | ړ | R | ARABIC LETTER REH WITH RING |
| 0x696 | ږ | 2 | ARABIC LETTER REH WITH DOT BELOW AND DOT ABOVE |
| 0x632 | ز | z | ARABIC LETTER ZAIN (ZAI) |
| 0x698 | ژ | 3 | ARABIC LETTER JEH |
| 0x633 | س | s | ARABIC LETTER SEEN |
| 0x634 | ش | \$ | ARABIC LETTER SHEEN |
| 0x69a | ښ | 5 | ARABIC LETTER SEEN WITH DOT BELOW AND DOT ABOVE |
| 0x635 | ص | S | ARABIC LETTER SAD |
| 0x636 | ض | D | ARABIC LETTER DAD |

| CODEPOINTS | Pashto | ROMAN | CODEPOINT DESCRIPTION |
|------------|--------|-------|------------------------------------|
| 0x637 | ط | T | ARABIC LETTER TAH |
| 0x638 | ظ | Z | ARABIC LETTER ZAH |
| 0x639 | ع | E | ARABIC LETTER AIN |
| 0x63a | غ | g | ARABIC LETTER GHAIN |
| 0x641 | ف | f | ARABIC LETTER FEH |
| 0x642 | ق | q | ARABIC LETTER QAF |
| 0x6a9 | ک | k | ARABIC LETTER KEHEH |
| 0x6ab | ک | G | ARABIC LETTER KAF WITH RING |
| 0x644 | ل | l | ARABIC LETTER LAM |
| 0x645 | م | m | ARABIC LETTER MEEM |
| 0x646 | ن | n | ARABIC LETTER NOON |
| 0x6bc | ن | N | ARABIC LETTER NOON WITH RING |
| 0x647 | ه | h | ARABIC LETTER HEH |
| 0x648 | و | w | ARABIC LETTER WAW |
| 0x6cc | ی | Y | ARABIC LETTER FARSI YEH |
| 0x64a | ي | y | ARABIC LETTER YEH |
| 0x6d0 | ې | e | ARABIC LETTER E |
| 0x6cd | ی | l | ARABIC LETTER YEH WITH TAIL |
| 0x626 | ئ | } | ARABIC LETTER YEH WITH HAMZA ABOVE |
| 0x621 | ء | ` | ARABIC LETTER HAMZA |
| 0x64b | ټ | F | ARABIC FATHATAN |

Table 3: Pashto Characters and Codepoints

Short vowels are not written in Pashto. Like all Arabic scripted languages the short vowels (written as diacritics) are only utilized for specific pronunciation purposes. Therefore vowel diacritics are not used at any stage of the orthographic transcription. The pronunciation lexicon delivered with this project was created from a wordlist generated from the orthographic transcription. As part of our internal processes each word in the wordlist was vowelized to capture the full pronunciation of each syllable. Vowelization of the wordlist utilized a set of specialized vowel diacritics, as listed in Table 4 below.

| CODEPOINTS | Pashto | ROMAN | CODEPOINT DESCRIPTION |
|------------|--------|-------|-----------------------|
| 0x64e | َ | a | ARABIC FATHA |
| 0x650 | ِ | i | ARABIC KASRA |
| 0x64f | ُ | u | ARABIC DAMMA |
| 0x652 | َ | o | ARABIC SUKUN |
| 0x64d | ټ | @ | ARABIC KASRATAN |

Table 4: Pashto vowelization diacritics

4. Description of character set used for orthographic transcription

Pashto is written in an extended Arabic script. The script consists of 27 standard Arabic script letters (plus Alef Madda), plus 17 symbols from an extended Arabic set. These additional symbols are needed for the so-called Persian letters, including /p/, /tS/ and a dotless /y/, as well as 14 symbols characteristic of the Pashto orthography and phonology: 6 retroflex consonants /t̪/, /d̪/, /s̪/ /z̪/ /n̪/ and /r̪/; 3 affricate and fricatives /dz/, /ts/ and /Z/; a ringed kaf (pronounced as /g/); and 3 additional /y/ symbols.

Because Pashto has borrowed many Arabic words and often maintains the original Arabic spelling despite pronunciation changes in Pashto, several Arabic letters map to single sounds – for example, those symbols which correspond to the sounds written in SAMPA as /s/, /z/, and /t/ (see Table 6 below).

The Zero-Width-Non-Joiner (ZWNJ) character is a non-printing character that is included in the Pashto database. The Unicode is U+200C. This character is placed between two characters that would otherwise be connected into a ligature, to ensure that they are displayed in their final and initial forms, respectively. For example, the character ‘h’ is always followed by a ZWNJ. The ZWNJ acts in a similar way as space, but keeps the characters on either side closer together. ZWNJ character is not used as an indicator of word boundaries.

5. Description of Romanization scheme

A modified Buckwalter Romanization scheme was adopted for this project; see Table 5. Please note that these are just the symbols required for the base lexicon (Kandahari variety) from which the four dialectal varieties are made.

| PASHTO | ROMAN | SAMPA | PASHTO | ROMAN | SAMPA |
|--------|-------|-------|--------|-------|-------|
| ا | A | A | ع | E | ? |
| آ | | A | غ | g | G |
| ب | b | b | ف | f | f |
| پ | P | p | ق | q | q |
| ت | t | t | ک | k | k |
| ښه | 8 | t' | گ | G | g |
| ص | v | s | ل | l | l |
| ج | j | dZ | م | m | m |
| خ | 7 | dz | ن | n | n |
| چ | J | tS | ڼ | N | n' |
| غ | 6 | ts | ه | h | h |
| ح | H | h | و | w | w |
| خ | x | x | ی | Y | j |
| د | d | d | ي | y | j |
| ډ | 9 | d' | ئ | e | e |
| ذ | ^ | z | ی | I | j |
| ر | r | 4 | ئ | } | j |
| ږ | R | r' | ا | a | a |
| ښ | 2 | z' | ي | i | i |
| ز | z | z | و | u | u |
| ژ | 3 | Z | و | o | + |
| س | s | s | ف | F | n |
| س | \$ | S | @ | @ | @ |
| س | 5 | s' | ء | ` | ? |
| س | S | s | = | = | _ |
| س | D | z | | | " |
| س | T | t | | | |
| س | Z | z | | | |

Table 5: Romanization Scheme of Afghan Pashto (Kandahari variety)

6. Description of method for word boundary detection

Word boundaries in the orthography are determined by location of white spaces.

7. Table containing all phonemes in the stipulated notation

| Pashto | IPA | X-SAMPA | Pashto | IPA | X-SAMPA |
|--------|-----------------|-----------------|--------|-------------------------|-----------------|
| ب | b | b | ض | z | z |
| پ | p | p | ث س ص | s | s |
| د | d | d | ژ | ʒ / dʒ / ʒ / z | Z / dZ / Z / z |
| ت ط | t | t | ش | ʃ | S |
| ډ | ɖ | d` | ږ | ʒ / g / ʒ / j | z` / g / Z / j\ |
| ټ | t̪ | t` | ښ | ʂ / x / ʃ / ç | s` / x / S / C |
| ج | dʒ | dZ | ف | f | f |
| چ | tʃ | tS | ح ه | h | h |
| خ | ɖz / z / dʒ / z | dz / z / dz / z | غ | ɣ | G |
| څ | ts / s / ts / s | ts / s / ts / s | خ | x | x |
| ق | q/k | q / k | و | w / o | w / o |
| ک | k | k | ی ی | j | j / i |
| ګ | g | g | ی ی | j | j - offglide |
| ع | (ʔ) | (ʔ) | ا | a | A |
| م | m | m | ې | e | e |
| ن | n | n | اَ | a | a |
| ن | ɳ | (N) | اِ | ɪ | i i: |
| ن | ɳ | n` | اُ | u | u u: |
| ل | l | l | اَ | ə | @ |
| ر | ɽ | r` | اَ | (no vowel) | (+) |
| ر | r | 4 | | (primary/strong stress) | “ |
| ذ ز ظ | z | z | | | |

Table 6: Pashto Phone Chart

Please note that in Table 6 alternative IPA and SAMPA symbols for some consonants (shaded in yellow) represent variation across the following four dialects respectively: South-West (e.g., Kandahar, the written standard), North-East (e.g., Peshawar), South-East (e.g., Quetta) and Central (least common variants, e.g., Central Ghilzai, Wardak, some outlying Kabuli). Alternative Pashto characters with a single IPA or SAMPA symbol represent a collapsing of an Arabic distinction to a single Pashto phoneme. Variation in the semivowel symbols, /w/ and any of the /j/ sounds represent their range of functions as consonants, off-glides and vowels in Pashto, according to context. There is no definitive IPA or X-SAMPA symbol for /ɽ/ the Pashto "retroflex lateral flap", so we use the symbols for "retroflex flap": ɽ and r`.

Pashto phonology is fairly complex, with both alveolar and retroflex stops and nasals. The most characteristic feature of the orthography is the presence of five *yah* symbols, which leads to some spelling variation. There is a large amount of dialectal variability in Pashto phonology, which has not yet been completely or systematically described in the literature. Table 6 presents the combined phoneme inventory for the four main Pashto dialects using the International Phonetic Alphabet (IPA) and the SAMPA keyboard equivalents. More specific observations relating to vowels, consonants, stress and phonological processes are given below.

Vowels

The following vowel classification has been used for phonetic representations in constructing the Pashto lexicon. There are 7 short vowels /i e @ a u o A/; there is a perceivable, if not contrastive, distinction between the short vowels / i u / and the long variants / i: u:/, although the short vowel /i/ generally neutralizes with the vowel /e/; All dialects have a schwa-like vowel as a distinct vowel phoneme which can sustain stress.

Consonants

- Pashto has an affinity for initial consonant clusters of two or three consonants. There are many consonant combinations and in some dialects, according to the literature, they may undergo metathesis, assimilation and dissimilation. e.g.,

| | |
|------|-------------|
| خبښ | x a s` |
| خبښ | s` a x |
| نسخه | n u . s x a |
| نسخه | n u x . s a |

However, metathesis has not been widely attested in the present data collection.

- There are a number of consonants only used in pronouncing loan words or words from Arabic. e.g., the pharyngeals /q/ and /ħ /h/. Their use in loan words indicates a more elegant speech. These pronunciations are dropped or changed to other sounds in general speech.

Stress

Stress is variable and distinctive in Pashto. Stress is used in verbs to indicate tense and mood. Stress is not indicated in the typical orthography but some dictionaries indicate the strong stress position. In this database primary stress is coded for all words in the pronunciation lexicon. The stress symbol precedes the stressed syllable.

Phonological processes

The following is a list (not exhaustive) of some known phonological processes and contextual pronunciation variants in Pashto. Dialect region is indicated where the process is known to be specific to a particular region.

• South West/East

- De-palatalization is common in the south west and south east dialects where /S/ > [s].
- /s`/ may elide in some consonant clusters in the South West dialect.
- /z/ may palatalize > [Z] in some words in the South East dialects.
- /s`/ may > [S] in some words in the South East dialects.

• North West/East

- /ts/ > [tS] and /dz/ > [dZ] in the North West dialects.
- /Z/ de-palatalizes > [z] in some cases in the North East dialects.
- /z`/ > [dZ] in some cases in the North East dialects.
- /s`/ and /z`/ have a variable realization in the North West dialects.
- /s`/ > [S] or [x]; /z`/ > [Z] or [g].
- Metathesis occurs e.g., "foot" /ps`a/ > [xpa] or [Spa].

Other processes

- (i) /n/ > [n`] when it precedes the (homorganic) alveolar retroflex stop.
- (ii) /h/ is frequently dropped when it occurs word-finally after a vowel.

Foreign loan words

There are a large number of loan words in Pashto. Speakers vary in the degree to which they approximate the original sounds in these words. Loan words come from languages including Arabic, Persian (Farsi and Dari), Urdu, Punjabi, English and Greek.

7.1. Complete List of all rare phonemes

No lists of rare phonemes were provided.

8. Other language items

8.1. Table of digits

There are three possible number systems to use in Afghan Pashto. Reading material contains only the Latin Digits (as in the leftmost column below).

| Digit | Pashto |
|-------|--------|
| 0 | ۰ |
| 1 | ۱ |
| 2 | ۲ |
| 3 | ۳ |
| 4 | ۴ / ۴ |
| 5 | ۵ / ۵ |
| 6 | ۶ / ۶ |
| 7 | ۷ |
| 8 | ۸ |
| 9 | ۹ |

Table 7: Digits

8.2 Other numbers

| Number | Pashto | Romanization |
|--------|--------|--------------|
| zero | صفر | Sfr |
| one | يو | Yw |
| two | دوه | Dwh |
| three | درې | Dre |
| four | څلور | 6lwr |

| Number | Pashto | Romanization |
|----------------------|------------|--------------|
| five | پښخه | Pen7h |
| six | شپږ | \$P2 |
| seven | اووه | Awwh |
| eight | اته | Ath |
| nine | نهه | nhh |
| ten | لس | ls |
| eleven | يوولس | Ywwls |
| twelve | دوولس | dwwls |
| thirteen | ديارلس | dYArIs |
| fourteen | څوارلس | 6wArIs |
| fifteen | پښخلس | Pen7Is |
| sixteen | شپاړس | \$PARs |
| seventeen | اوولس | AwwIs |
| eighteen | اتلس | AtIs |
| nineteen | نولس | nwIs |
| twenty | شل | \$I |
| thirty | دېرش | der\$ |
| forty | څلوېښت | 6lwe5t |
| fifty | پښخوس | Pen7ws |
| sixty | شپيته | \$Peth |
| seventy | اويا | AwYA |
| eighty | اتيا | AtYA |
| ninety | نوي | nwy |
| hundred | سل | sl |
| thousand | زر | zr |
| one hundred thousand | يو لک | Ywlk |
| million | ميليون | milYwn |
| ten million | لس ميليونه | ls milYwnh |

Table 8: Numbers

8.3 Spelling alphabet - Pashto symbols

Table 9 below provides the X-SAMPA pronunciation of English letters and digits as found in initialisms borrowed from English (e.g., DVD) as attested in the corpus.²

| English | Pashto | Romanization | X-SAMPA | English | Pashto | Romanization | X-SAMPA |
|---------|--------|--------------|---------|---------|--------|--------------|-----------------|
| A | اي | Ae | e | S | اېس | Aeso | e s |
| B | بي | biy | b i: | T | تي | 8iy | t` i: |
| C | سي | siy | s i: | U | يو | Yuw | j u: |
| D | دي | 9iy | d` i: | V | وي | wiy | w i: |
| E | اي | Ay | i: | W | ډبليو | 9aboloYuw | d` a b . l j u: |
| F | اېف | Aefo | e f | X | اېکس | Aekoso | e k s |
| G | جې | jiy | dZ i: | Y | N/A | N/A | N/A |
| H | اېچ | AeJo | e tS | Z | N/A | N/A | N/A |
| I | اي | Yo | A j | 0 | N/A | N/A | N/A |
| J | جې | je | dZ e | 1 | ون | wano | w a n |
| K | کې | ke | k e | 2 | ټو | 8uw | t` u: |
| L | اېل | Aelo | e l | 3 | تري | toriy | t 4 i: |
| M | اېم | Aemo | e m | 4 | فور | fwro | f o 4 |
| N | اېن | Aeno | e n | 5 | N/A | N/A | N/A |
| O | او | Awo | o | 6 | N/A | N/A | N/A |
| P | پې | Piy | p i: | 7 | N/A | N/A | N/A |
| Q | N/A | N/A | N/A | 8 | N/A | N/A | N/A |
| R | آر | Iro | A 4 | 9 | N/A | N/A | N/A |

Table 9: Extant initialisms from the corpus

The following table provides X-SAMPA for Perso-Arabic names of Pashto letters. Note that these pronunciations have not been attested in the corpus. However, it represents the way native words *could* be spelled out in theory.

| Letter | IPA | Pashto Name in X-SAMPA |
|--------|------|------------------------|
| ا | -- | Allf |
| آ | ɑ | Allf mad |
| أ | an | Allf tanwln |
| ب | b | be |
| پ | p | pe |
| ت | t | te |
| ټ | t̚ | t` e |
| ث | s | se |
| ج | dʒ | dZlm |
| چ | tʃ | tSlm/tSe |
| خ | dʒ/z | dze/dZlm/ze |
| ځ | tʃ/s | tse/tslm/se |
| ح | ø/h | he/halwa he |
| خ | x | xe |
| د | d̪ | dAl |
| ډ | d̪ | d` al |
| ذ | z | zAl |
| ر | r | 4e |

² N/A = not attested in corpus.

| Letter | IPA | Pashto Name in X-SAMPA |
|--------|---------|----------------------------|
| ر | ɽ | r`e |
| ز | z | ze |
| ژ | ʒ/dʒ/z | Ze |
| ډ | z/ʒ/ɟ/g | ze/ge |
| س | s | sIn |
| ش | ʃ | SIn |
| ښ | ʂ/x | s`In/xIn |
| ص | s | sAd/swAd/sxwAt/swA |
| ض | z | zAd/zwAd/zxwAt/zwA |
| ط | t̪ | tA/twe/txwe/toj/twA |
| ظ | z | zA/zwe/zGwe/zoj/zGA |
| ع | ʔ/ʕ | ?ajn/ajn |
| غ | ɣ | Gajn |
| ف | f | fe |
| ق | k | qAf/qaf |
| ک | k | kAf/kaf |
| گ | g | gAf |
| ل | l | lAm |
| م | m | mim |
| ن | n | nun |
| ڼ | ɳ | n`un |
| و | w/o/u | wAw |
| ؤ | ø/w | wAw-hamza |
| ه | h | he/hA |
| ۀ | ə | he-hamza/zwarakaj |
| ی | aj | ye/yA/mulajana he/prata je |
| ي | i/y | saxt@ je |
| ې | e | pasta je |
| ى | ej | s@dzina je |
| ئ | ej | k@4wAla je |

Table 10: Pronunciation of Pashto letters

9. References

- Elfenbein, Josef. "Pashto Phonology", *Phonologies of Asia and Africa* (Indiana: Eisenbrauns, 1997).
<https://www.cia.gov/library/publications/the-world-factbook/geos/pk.html>
 International Communication Union <http://www.itu.int/oth/T0202000001/en>
 International Communication Union <http://www.itu.int/oth/T02020000A1/en>
 Vonwiller J., Bishop J., "The Pashto Language Summary" (Appen Pty Limited) 2005.