



The
Word Order in
Western Asia
Corpus

metadata for the data set

Iranian — Persian (New)

11 August 2022
(docv1.1)

Overview of the data set

data set	iran_persian_new
contributors	Elham Izadi
published	31 July 2022
texts	30
tokens	1628 analyzed (3219 total)
notes	XLS/TSV contain entire source text

Metadata for individual texts

Text A

source	Haig & Rasekh-Mahand (2022: ac_m_corona1) (WAV ✓ PDF ✓)
location	exact location unknown (34.7983° N, 48.5150° E)
recorded	2020
text type	podcast
speaker	01 — male, age 45
tokens	46 analyzed (91 total)

Text B

source	Haig & Rasekh-Mahand (2022: ac_m_corona2) (WAV ✓ PDF ✓)
location	exact location unknown (34.7983° N, 48.5150° E)
recorded	2020
text type	podcast
speaker	01 — male, age 45
tokens	63 analyzed (130 total)

Text C

source	Haig & Rasekh-Mahand (2022: ac_m_depression) (WAV ✓ PDF ✓)
location	exact location unknown (34.7983° N, 48.5150° E)
recorded	2020
text type	radio interview
speaker	02 — male, age 35
tokens	84 analyzed (126 total)

Text D

source Haig & Rasekh-Mahand (2022: acd_m_plane) (WAV ✓ PDF ✓)
location exact location unknown (34.7983° N, 48.5150° E)
recorded 2020
text type radio interview
speaker 03 — male, age 70
tokens 47 analyzed (135 total)

Text E

source Haig & Rasekh-Mahand (2022: acd_m_ufo) (WAV ✓ PDF ✓)
location exact location unknown (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 03 — male, age 70
tokens 33 analyzed (92 total)

Text F

source Haig & Rasekh-Mahand (2022: oh_f_accident) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2017
text type personal narrative
speaker 04 — female, age 50
tokens 124 analyzed (279 total)

Text G

source Haig & Rasekh-Mahand (2022: oh_f_amirali) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 05 — female, age 34
tokens 27 analyzed (50 total)

Text H

source Haig & Rasekh-Mahand (2022: oh_f_aunt) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2018
text type personal narrative
speaker 06 — female, age 50
tokens 65 analyzed (172 total)

Text I

source Haig & Rasekh-Mahand (2022: oh_f_childhood1) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2018
text type personal narrative
speaker 06 — female, age 50
tokens 50 analyzed (135 total)

Text J

source Haig & Rasekh-Mahand (2022: oh_f_childhood2) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 07 — female, age 39
tokens 42 analyzed (94 total)

Text K

source Haig & Rasekh-Mahand (2022: oh_f_class) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2019
text type personal narrative
speaker 08 — female, age 43
tokens 51 analyzed (90 total)

Text L

source Haig & Rasekh-Mahand (2022: oh_f_daryush) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2019
text type personal narrative
speaker 07 — female, age 38
tokens 52 analyzed (92 total)

Text M

source Haig & Rasekh-Mahand (2022: oh_f_istanbul1) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2022
text type personal narrative
speaker 09 — female, age 29
tokens 42 analyzed (70 total)

Text N

source Haig & Rasekh-Mahand (2022: oh_f_marry) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2018
text type personal narrative
speaker 06 — female, age 50
tokens 99 analyzed (170 total)

Text O

source Haig & Rasekh-Mahand (2022: oh_f_mask) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 10 — female, age 31
tokens 25 analyzed (40 total)

Text P

source Haig & Rasekh-Mahand (2022: oh_f_parham) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 11 — female, age 34
tokens 71 analyzed (112 total)

Text Q

source Haig & Rasekh-Mahand (2022: oh_f_pool) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 12 — female, age 35
tokens 35 analyzed (61 total)

Text R

source Haig & Rasekh-Mahand (2022: oh_f_taxi1) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 13 — female, age 42
tokens 71 analyzed (131 total)

Text S

source Haig & Rasekh-Mahand (2022: oh_f_uncle1) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2019
text type personal narrative
speaker 14 — female, age 43
tokens 31 analyzed (77 total)

Text T

source Haig & Rasekh-Mahand (2022: oh_f_uncle2) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2018
text type personal narrative
speaker 06 — female, age 50
tokens 63 analyzed (116 total)

Text U

source Haig & Rasekh-Mahand (2022: oh_f_university1) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 15 — female, age 33
tokens 42 analyzed (79 total)

Text V

source Haig & Rasekh-Mahand (2022: oh_m_corona3) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 16 — male, age 39
tokens 14 analyzed (37 total)

Text W

source Haig & Rasekh-Mahand (2022: oh_m_military1) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 17 — male, age 60
tokens 114 analyzed (202 total)

Text X

source Haig & Rasekh-Mahand (2022: oh_m_military2) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 18 — male, age 34
tokens 33 analyzed (59 total)

Text Y

source Haig & Rasekh-Mahand (2022: oh_m_music) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2019
text type personal narrative
speaker 19 — male, age 45
tokens 113 analyzed (185 total)

Text Z

source Haig & Rasekh-Mahand (2022: oh_m_taxi2) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 20 — male, age 32
tokens 45 analyzed (103 total)

Text ZA

source Haig & Rasekh-Mahand (2022: oh_m_television) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 21 — male, age 29
tokens 30 analyzed (57 total)

Text ZB

source Haig & Rasekh-Mahand (2022: oh_m_university2) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 22 — male, age 60
tokens 37 analyzed (69 total)

Text ZC

source Haig & Rasekh-Mahand (2022: oh_m_university3) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 23 — male, age 60
tokens 51 analyzed (104 total)

Text ZD

source Haig & Rasekh-Mahand (2022: oh_m_usa) (WAV ✓ PDF ✓)
location Hamedan, Iran (34.7983° N, 48.5150° E)
recorded 2020
text type personal narrative
speaker 24 — male, age 31
tokens 28 analyzed (61 total)

Background to the data set

The texts in this data set stem from the HamBam corpus (*Hamedan-Bamberg Corpus of Contemporary Spoken Persian*, Haig & Rasekh-Mahand 2022) and represent colloquial spoken Persian as spoken by speakers living in and around the city of Hamedan in Iran. Texts F, H, I, N, T, W, Y, ZB, and ZC are closer to the colloquial variety of Hamedan, the others closer to the colloquial variety of Tehran. See Haig & Rasekh-Mahand (2022) for more details of the HamBam corpus. Note that the age of the speakers has been estimated in many cases, and is thus not an exact value.

References

Haig, Geoffrey & Rasekh-Mahand, Mohammad (eds.). 2022. *HamBam: Hamedan-Bamberg Corpus of Contemporary Spoken Persian*. (<https://multicast.aspra.uni-bamberg.de/resources/hambam/>) (Accessed 2022-07-24).