Multi-CAST

Multilingual Corpus of Annotated Spoken Texts

Persian

— metadata sheet —

July 2016

Shirin Adibifar

University of Bamberg

Language Archive Cologne

Citing the corpus

The Multi-CAST Persian corpus should be cited as follows:

```
Shirin Adibifar. 2016. Persian.
 In Haig, Geoffrey & Schnell, Stefan (eds.),
 Multi-CAST (Multilingual Corpus of Annotated Spoken Texts).
 (https://lac.uni-koeln.de/en/multicast-persian/)
 (date accessed.)
```

Language data

language	Persian (ISO 639-3: pes)
affiliation	Indo-European, Iranian, Southwestern
area spoken	Iran
varieties recorded	Farsi, Tehran / Sari
text type	stimulus-based narratives

Corpus data

file name	speaker	date rec'd	text type	length h:mm:ss	clause units
persian_g1-f-01	PS01	2015	stim.-based	1:34	47
persian_g1-f-02	PS02	2015	stim.-based	2:10	54
persian_g1-f-05	PS05	2015	stim.-based	2:15	60
persian_g1-f-07	PS07	2015	stim.-based	1:05	38
persian_g1-f-08	PS08	2015	stim.-based	1:40	25
persian_g1-f-09	PS09	2015	stim.-based	4:29	100
persian_g1-f-10	PS10	2015	stim.-based	3:19	83
persian_g1-f-11	PS11	2015	stim.-based	1:42	60
persian_g1-f-12	PS12	2015	stim.-based	1:45	49
persian_g1-f-14	PS14	2015	stim.-based	3:03	99
persian_g1-m-03	PS03	2015	stim.-based	0:45	17
persian_g1-m-04	PS04	2015	stim.-based	2:03	61
persian_g1-m-06	PS06	2015	stim.-based	0:51	22
persian_g1-m-13	PS13	2015	stim.-based	2:50	69
persian_g2-f-01	PS15	2015	stim.-based	2:23	58
persian_g2-f-02	PS16	2015	stim.-based	1:27	44
persian_g2-f-03	PS17	2015	stim.-based	1:37	40
persian_g2-f-04	PS18	2015	stim.-based	1:03	25
persian_g2-f-05	PS19	2015	stim.-based	1:52	26
persian_g2-f-06	PS20	2015	stim.-based	1:27	56
persian_g2-f-07	PS21	2015	stim.-based	1:41	51

⇒ *table continued from previous page*

file name	speaker	date rec'd	text type	length h:mm:ss	clause units
persian_g2-m-08	PS22	2015	stim.-based	1:44	49
persian_g2-m-09	PS23	2015	stim.-based	1:20	42
persian_g2-m-10	PS24	2015	stim.-based	1:17	41
persian_g2-m-11	PS25	2015	stim.-based	1:01	25
persian_g2-m-12	PS26	2015	stim.-based	1:08	40
persian_g2-m-13	PS27	2015	stim.-based	1:23	52
persian_g2-m-14	PS28	2015	stim.-based	1:03	36
persian_g2-m-15	PS29	2015	stim.-based	2:35	48
			totals for the Persian corpus	0:52:32	1,417

Table 1. List of files in the Persian corpus.

Background to the recordings

The Multi-CAST *persian* corpus consists of 29 texts, each from a different native speaker of Persian, recorded in 2014 and 2015. Each text was produced in an interview-like setting, in which the corpus compiler (Adibifar) showed the speaker a 6-minute video (the *Pear Story*, cf. Chafe 1980) on a Dell color laptop computer with a 14-inch screen. At the end of the film, the interviewees were asked to recount the actions of the film in their own words; the instructions were given by the researcher in their native language (Persian), and each participant received the same set of instructions. The interval between watching the movie and recounting its contents was less than 5 minutes. The participants were also were requested to provide basic information regarding their age, gender, level of education, places of socialization, languages of communication (inside and outside of the domestic setting), and language of their parents, as well as their contact information in case of further questions.

The first half of the recordings (*g1* in the text name) took place in a relaxed domestic setting in interviewer's hometown in Mazanderan province in the north of Iran, and in three cases, in the speakers' apartments in Tuebingen, Germany. The remainder (*g2* in the text name) were conducted with students from the Islamic Azad University in Tehran and the Behšahr University in Mazanderan Province in seminar rooms of the respective universities.

References

Chafe, Wallace (ed.). 1980. *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

speaker	gender	age	date born	date rec'd	clause units	texts
PS01	female	39	1976	2015	47	*persian_g1-f-01*
PS02	female	29	1986	2015	54	*persian_g1-f-02*
PS03	male	22	1993	2015	17	*persian_g1-m-03*
PS04	male	25	1990	2015	61	*persian_g1-m-04*
PS05	female	26	1989	2015	60	*persian_g1-f-05*
PS06	male	32	1983	2015	22	*persian_g1-m-06*
PS07	female	25	1990	2015	38	*persian_g1-f-07*
PS08	female	25	1990	2015	25	*persian_g1-f-08*
PS09	female	25	1990	2015	100	*persian_g1-f-09*
PS10	female	31	1984	2015	83	*persian_g1-f-10*
PS11	female	33	1982	2015	60	*persian_g1-f-11*
PS12	female	33	1982	2015	49	*persian_g1-f-12*
PS13	male	35	1980	2015	69	*persian_g1-m-13*
PS14	female	29	1986	2015	99	*persian_g1-f-14*
PS15	female	20	1995	2015	58	*persian_g2-f-01*
PS16	female	20	1995	2015	44	*persian_g2-f-02*
PS17	female	20	1995	2015	40	*persian_g2-f-03*
PS18	female	20	1995	2015	25	*persian_g2-f-04*
PS19	female	21	1995	2015	26	*persian_g2-f-05*
PS20	female	38	1977	2015	56	*persian_g2-f-06*
PS21	female	33	1982	2015	51	*persian_g2-f-07*
PS22	male	20	1995	2015	49	*persian_g2-m-08*
PS23	male	22	1993	2015	42	*persian_g2-m-09*
PS24	male	20	1995	2015	41	*persian_g2-m-10*
PS25	male	25	1990	2015	25	*persian_g2-m-11*
PS26	male	20	1995	2015	40	*persian_g2-m-12*
PS27	male	20	1995	2015	52	*persian_g2-m-13*
PS28	male	20	1995	2015	36	*persian_g2-m-14*
PS29	male	27	1988	2015	48	*persian_g2-m-15*

Table 2. List of Persian corpus speakers.

Corpus changelog

date	changes	affected texts		
2016–06	added 29 corpus files	*persian_**		
		g1-f-01,	*g1-f-02,*	*g1-f-05,*
		g1-f-07,	*g1-f-08,*	*g1-f-09,*
		g1-f-10,	*g1-f-11,*	*g1-f-12,*
		g1-f-14,	*g1-m-03,*	*g1-m-04,*
		g1-m-06,	*g1-m-13,*	*g2-f-01,*
		g2-f-02,	*g2-f-03,*	*g2-f-04,*
		g2-f-05,	*g2-f-06,*	*g2-f-07,*
		g2-m-08,	*g2-m-09,*	*g2-m-10,*
		g2-m-11,	*g2-m-12,*	*g2-m-13,*
		g2-m-14,	*g2-m-15*	