

# Turkmen

A Turkic language. Except when it's Russian.

Vivian Robison

# Demographics

- ~75,000 speakers, primarily in Turkmenistan, Iran, & Afghanistan
- Official language of Turkmenistan



# Notable Characteristics

- Vowel harmony
  - Primary: front vs. back
  - Secondary: rounded vs. unrounded (with exceptions)
  - /ɑ:/ and /æ:/ don't have rounded counterparts → exempt from secondary harmony AND block it from applying to subsequent syllables!
- Agglutination
  - one suffix can have 20 surface forms, minimum 2 if it has a vowel
  - lots of cross-boundary effects--syllable deletion, vowel deletion, consonant deletion, vowel length changes, assimilation, etc.

# Orthography Challenges



- Pre-1928: Arabic script
- 1928-1940: Latin script
- 1940-1991: Cyrillic script

Хемме адамлар өз мертебеси ве хукуклары бюнча дең ягдайда дүнйә инйәрлер. Олара аң хем выждан берлендир ве олар бир-бирлери билен доганлык рухундакы гарайышда болмалыдырлар.

- 1991-present: Latin script

Hemme adamlar öz mertebesi we hukuklary boýunça deň ýagdaýda dünýä inyärler. Olara aň hem wyždan berlendir we olar bir-birleri bilen doganlyk ruhundaky garaýyşda bolmalydyrlar.

# Resources?

- Parallel corpora exist--but are relatively small
- No available morphological analyzers specifically for Turkmen

# Previous MT work

- Turkmen→Turkish
  - Istanbul Technical University & Sabanci University (2006)
- Turkmen↔Turkish
  - Karadeniz Technical University (2007)
- There are multiple free Turkmen↔Russian automatic translators online of dubious quality

Baisa, V., V. Suchomel. “Large Corpora for Turkic Languages and Unsupervised Morphological Analysis.” *In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey.

[http://nlp.fi.muni.cz/~xsuchom2/papers/BaisaSuchomel\\_TurkicResources.pdf](http://nlp.fi.muni.cz/~xsuchom2/papers/BaisaSuchomel_TurkicResources.pdf)

Köse, C., G Amanmyradov, Ö. Özyurt, 2007. “A machine Translation System for Turkish-Turkmen Chat” *Journal of science and Technology* 1(1) 2007.

<http://ozcanozyurt.teknoogretim.com/yayinlar/bu-jst.pdf>

Tantuğ A.C., Adalı E., Oflazer K., 2007. “Machine Translation between Turkic Languages”, *In Proceedings of ACL 2007 – Companion Volume*, Prague, Czech Republic. <http://www.aclweb.org/anthology/P07-2048>

Turkmen. (n.d.). Retrieved from <http://www.ethnologue.com/language/tuk>

Turkmen. (n.d.). Retrieved from <http://multitree.linguistlist.org/codes/tuk>

with thanks to Nurjahan Orazgylyjova, my language consultant