

11-712: NLP Lab Report

Maya Tydykov

April 25, 2014

1 Basic Information about Russian

The Russian language is an Indo-European language spoken primarily in Russia and in other parts of the world by approximately 162 million people. It belongs to the Eastern branch of the Slavic language family (Lewis et al., 2013). Russian is a free word order language, although according to (Dryer and Haspelmath, 2013), it is primarily Subject-Verb-Object (SVO). Russian has three genders and six cases, which are marked via suffixes on words. It is written using the Cyrillic script, which was originally created for 9th-10th century Slavic language speakers in order to translate the Bible along with other church texts (“Cyrillic alphabet”).

2 Past Work on the Syntax of Russian

3 Available Resources

4 Survey of Phenomena in Russian

5 Initial Design

6 System Analysis on Corpus A

7 Lessons Learned and Revised Design

8 System Analysis on Corpus B

9 Final Revisions

10 Future Work

References

- Cyrillic alphabet. Encyclopaedia Britannica Online Academic Edition. URL <http://www.britannica.com/EBchecked/topic/148713/Cyrillic-alphabet>.
- Matthew S. Dryer and Martin Haspelmath, editors. *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013. URL http://wals.info/languoid/lect/wals_code_rus.
- M. Paul Lewis, Gary F. Simons, and Charles D. Fennig. *Ethnologue: Languages of the world*, seventeenth edition. Web, 2013. URL <http://www.ethnologue.com/statistics/size>.