Language in 10 minutes

Russian

Morgan Ulinski
Rehman Khan

Introduction

- Russian Language is a complex and diverse language with several dialects
- Russian MT involves several challenges
- Lot of work has been done already in Russian MT in the middle to late of the 20th Century.
- An urban legend involving a pair of MT systems between Russian and English. The legend proclaims that once someone fed a English–Russian MT system the bible verse "The spirit is willing, but the flesh is weak." When back translating the sentence with the Russian– English system, the system returned "The vodka is good but the meat is rotten."
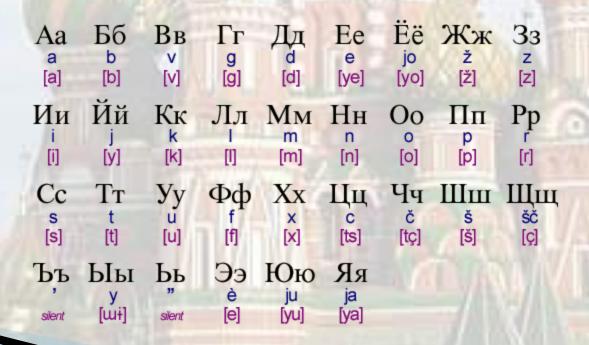
Russian (*russkiy)*

- Geographically widespread language
- Part of Slavic
  - East Slavic
    - Russian , Ukrainian
- Dialects
  - Northern Dialects
  - Southern Dialects
- Derived Languages
  - Several derived languages based on Russian also exist

Russian Alphabet

- Russian has 33 letters
- Written in modified cyrillic

Аа	Бб	Вв	Гг	Дд	Ее	Ёё	Жж	Зз
a	b	v	g	d	e	jo	ž	z
[a]	[b]	[v]	[g]	[d]	[ye]	[yo]	[ž]	[z]

Ии	Йй	Кк	Лл	Мм	Нн	Оо	Пп	Рр
i	j	k	l	m	n	o	p	r
[i]	[y]	[k]	[l]	[m]	[n]	[o]	[p]	[r]

Сс	Тт	Уу	Фф	Хх	Цц	Чч	Шш	Щщ
s	t	u	f	x	c	č	š	šč
[s]	[t]	[u]	[f]	[x]	[ts]	[tç]	[š]	[ç]

Ъъ	Ыы	Ьь	Ээ	Юю	Яя
'	y	"	è	ju	ja
silent	[ɯɨ]	silent	[e]	[yu]	[ya]

Typical Phrases

Cyrillic	Transliteration	English Meaning
У меня хороший день	U menya horoshīi den	I am having a good day
Как тебя зовут	Kak tebya zovut	What is your name?
Меня зовут Дмитрий	Menya zovut Dmitrīi	My name is Dmitri
Я не говорю на русском	Ya ne govoru na Russki	I don't speak Russian
Обучение русскому не интересно	Obuchenie russkomu ne interesno	Learning Russian is not interesting

Variations

Cyrillic	Я хороший	Я чувствую себя хорошо
Transliteration	Ya **horoshii**	Ya chustvuiy sebya **horosho**
Word For Word	I Good	I feel myself good
English	I am good	I feel good

- Horosho
    - Means Good
    - used when no relation to a subject
- Horoshii
    - Means Good
    - used when there is relation to a subject

Morphology

- Highly synthetic morphology
  - Uses inflections rather than word order to express grammatical structure
- Nouns/adjectives
  - 6+ cases (nominative, genitive, dative, accusative, instrumental, prepositional, other incomplete/degenerate cases)
  - 2 numbers (singular and plural)
  - Obeys grammatical gender (masculine, feminine, neuter)
  - **стол** = 'table', **окно** = 'window'

Case	Masculine	Neuter	Case Ending
Nominative	стол	окн-о	Nothing or -o/e
Accusative	стол	окн-о	= Nom or Gen
Genitive	стол-а	окн-а	-а/я (or -у)
Dative	стол-у	окн-у	-у/ю
Prepositional	стол-е	окн-е	-е (or -у)
Instrumental	стол-ом	окн-ом	-ом/ем

Morphology

- Verbs
  - Three persons (First, Second and Third)
  - Two numbers (singular/plural)
  - Two simple tenses(present/future and past)
    - Morphemes specify other tenses
  - Suffixes indicate active/passive

First Person Singular (я)	−у	First Person Plural (мы)	−ём
Second Person Singular (ты)	−ёшь	Second Person Plural (вы)	−ёте
Third Person Singular (он она оно)	−ёт	Third Person Plural (они)	−ут

- Word formation
  - Uses prefixes, infixes, and suffixes to produce multiple derivatives of a single word

Syntax

- Basic word order
  - usually SVO in transitive clauses, but flexible due to morphological information
  - free word order in intransitive clauses
- Noun phrases do not have a definite/indefinite article; meaning is inferred from context
- Adjectives usually come before nouns
- Example
  - Красная книга находится на столе
  - Krasnaya kniga nahodit·sya na stole
  - The red book is on the table

Challenges in English To Russian Translation

- Use of auxiliary verbs to represent tenses
- Russian's lack of articles can cause difficulties when translating into English
- Russian is morphologically rich, whereas English is not
- Complex morphology creates data sparsity problems

Machine Translation Research

- Most of the research work on Machine Translation was done in the 50s to 80s
- Nowadays Russian to non-English(Chinese etc) are active areas
- Good Readings
  - 1967 D.M. Yates
    - A computer model for Russian grammatical description, and a method of English synthesis in machine translation
  - 1987 Jan Hajic
    - RUSLAN: an MT system between closely related languages

Online Resources

- Lots of resources online
  - InterTran & NeuroTran http://tranexp.com/
  - Logomedia / LEC http://www.lec.com/default.asp
  - Lingvistica 98(Language Master) http://www.ling98.com/
  - Systran – http://www.systransoft.com/
  - PROMT – http://www.promt.ru/
  - WinBabel – http://www.winbabel.com/

Conclusion

- Lot of work has been done in Russian-English MT
- Some challenges still exist
- Lot of work still can be done to translate Russian into other languages
- Спасибо (Spasibo/Thank You) .....