



DLIFLC
DEFENSE LANGUAGE INSTITUTE
FOREIGN LANGUAGE CENTER

Linguistic Characteristics of Russian Listening Texts by ILR Levels

Dr Aleksey Novikov, Prof Irene Krasner

DISCLAIMER

This presentation has been approved for public release by the Defense Language Institute Foreign Language Center's Public Affairs Office. For verification please e-mail: mpao@dliflc.edu

Contents of this presentation are not necessarily the official views of the Defense Language Institute Foreign Language Center, nor are they endorsed by the Department of the Army, the Department of Defense, or the U.S. Government.

All third party products/materials featured in the presentation remain the intellectual property of their respective authors/originators. Use of outside materials is done under the fair use copyright principle, for educational purposes only.

The content of this presentation is the sole responsibility of the authors.

AGENDA

- METHODOLOGY OF TEXT COLLECTION
 - TEXT TYPOLOGY
- RESEARCH METHODOLOGY
 - RESEARCH QUESTIONS
- DATA ANALYSES
 - WORD FREQUENCY
 - MULTI-WORD EXPRESSIONS
 - SYNTAX AND MORPHOLOGY
- FINDINGS AND INTERPRETATIONS

TEXT FUNCTIONS

Level	Function	Examples
0+	Enumerate	List of items such as a shopping list
	Collect Information	As in a passport application form
1	Inform (Orientation)	About places and events, such as the location of a place or times
	Announce	Social or public events, such as a wedding or a meeting
	Advertise	Common products and services

TEXT FUNCTIONS

Level	Function	Examples
2	Direct	On how to get to places
	Report -Events	Current events
	Report -Studies	Studies and discoveries, such as a medical discovery
	Explain	How things and processes work, such as a new technological device
	Instruct	How to accomplish regular tasks such as a work assignment
	Narrate	Recounting of personal stories and common occurrences
	Describe	People and places

TEXT FUNCTIONS

Level	Function	Examples
3/4	Compare	Ideas or complex facts
	Comment	On ideas and events, such as in a letter to the editor
	Analyze	Abstract ideas and complex facts
	Advance Opinion	On a particular issue advocated by an author
	Counter Opinion	Opposition to an idea or opinion expressed by another party
	Hypothesize	On a past or future event
5	Advise	On social behaviors and courses of action

LEVEL 1

- Able to understand simple spoken language dealing with **basic information**, often with great difficulty and **in limited contexts**. Understands **general subject matter** in **simple spoken material**. Can understand straightforward instructions and simple statements and questions. Can understand explicitly stated simple messages in non-participatory listening when supported by redundancy or context. Understands connected phrases and sentences in **simple speech**. Understands **some high frequency vocabulary**. Understands **some basic structures**. May distinguish some different sounds and tones.
- Understanding is limited to self, family, or **immediate environment**, including, but not limited to, meals, time, directions, simple transactions, or personal information. Can identify general subject matter in **some simple speech**. Generally, understands short exchanges and statements. Able to understand speech that fills **basic courtesy requirements**. Not able to distinguish formality other than in **simple phrases and expressions**. Demonstrates an extremely limited understanding of cultural norms and expectations. Cannot easily understand mood, feeling, or emotion conveyed in speech. Understands **some spoken language** during participatory listening, particularly if delivered at a slower rate than normal, with frequent repetitions or paraphrase (<https://www.govtilr.org/Skills/Listening.htm>)

THE USE OF GENERIC WORDS TO DESCRIBE LEVELS

From generic descriptors to more specific fine-grained features representative of each level

- What is basic?
- What is simple?
- What is straightforward?
- What is normal?

RESEARCH QUESTIONS

1. How can we describe ILR SLD levels in Russian listening passages in terms of word frequency?
2. How can we describe ILR SLD levels in Russian listening passages in terms of multiword expressions?
3. How can we describe ILR SLD levels in Russian listening passages in terms of grammar (syntax and morphology)?

SOURCES OF TEXTS

Transcripts from authentic and semi-authentic audio sources

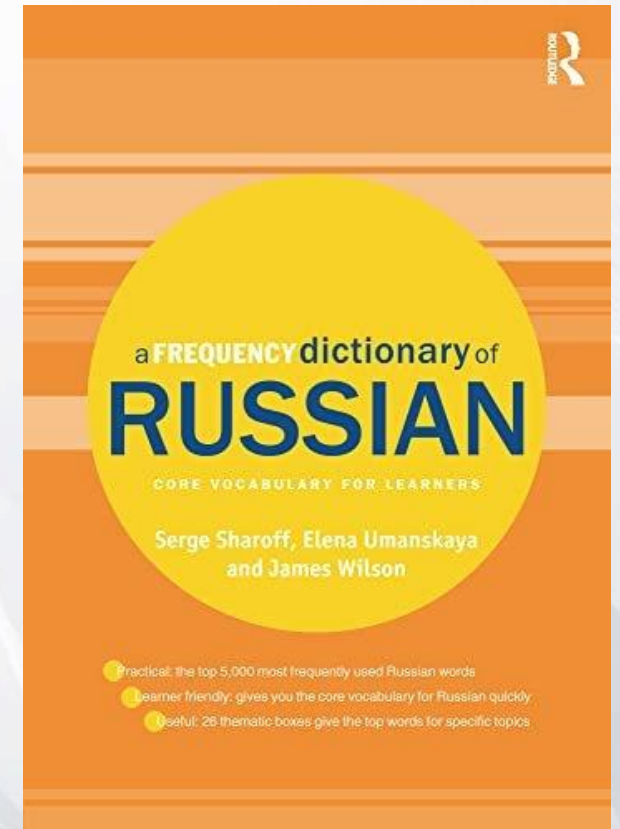
- Diagnostic Assessment Data bank
- NFLC texts
- Undergraduate School Russian program textbook
- Low-level listening project

DESCRIPTIVE STATS

ILR Level	Number of Texts	Mean of Words per Text	SD for the Number of Words per Text
Level 1	69	48	24
Level 2	68	151	52
Level 3	72	243	64

5,000 MOST FREQUENT WORDS?

- 4,000 -5,000 most frequent words account for about 95% of all written text, and 1,000 most frequent words account for about 85% of speech (Nation, 1990)
- Based on the Russian Internet Corpus (150 million Russian words based on more than 30,000 pages)
- Genres:
 - Reporting – 10%
 - Fiction – 27%
 - Legal texts – less than 1%
 - Instruction 7%
 - Discussion – 55%



TEXT MEAN FREQUENCY METHODOLOGY

1. Texts were tokenized and lemmatized using Spacy
2. The lemmas were assigned a frequency from the Russian Frequency Dictionary
3. A mean text frequency was calculated for each text
4. If words were not found in the Frequency Dictionary, they were assigned a zero
5. Grand mean was calculated for each level

TEXT MEAN FREQUENCY FINDINGS

ILR Level	Grand Mean Frequency	SD
Level 1	4928	1777
Level 2	4774	910
Level 3	4765	827

REMOVING STOP WORDS

ILR Level	Grand Mean Frequency (Without Stopwords)	SD
Level 1	284	129
Level 2	305	96
Level 3	332	84

Stopwords	English
а	But; and
без	Without
более	More
давно	A while ago
до	Before
если	if
некоторых	Some
тебя	you
...	

REMOVING A CUSTOM LIST

ILR Level	Grand Mean Frequency (without custom list)	SD
Level 1	661	278
Level 2	691	187
Level 3	841	143

Custom list

И (and)	Этот (this)
В (to; in)	По (along)
Не (not)	К (towards)
На (on; to)	Но (but)
Я (I)	Они (they)
Быть (to be)	Мы (we)
С (with; from)	Она (she)
Он (he)	Как (how; as)
А (but; and)	То (that)
Это (this)	Который (which)

EXAMPLES OF LEVEL 1 WORDS WITH THEIR FREQUENCIES

Level 1 words	Corresponding Frequency (Russian Frequency Dictionary)	English Translation
Сумка	50.04	Bag
Куртка	28.24	Jacket
Галстук	17.22	Tie
Скидка	17.36	Discount
Отличный	61.36	Excellent
Парк	58.37	Park
Кинотеатр	20.51	Movie theater
Тёплый	78.38	Warm
Ресторан	63.73	Restaurant
Снежный	22.4	Snowy

EXAMPLES OF LEVEL 2 WORDS WITH THEIR FREQUENCIES

Level 2 words	Corresponding Frequency (Russian Frequency Dictionary)	English Translation
Факт	189.96	Fact
Обстоятельство	105.46	Circumstance
Животное	102.22	Animal
Примерно	126.78	Approximately
Житель	116.72	Resident
Проверять	123.49	Check
Военный	217.83	Military
Ехать	131.84	Go by car
Мировой	151.52	World

EXAMPLES OF LEVEL 3 WORDS WITH THEIR FREQUENCIES

Level 3 words	Corresponding Frequency (Russian Frequency Dictionary)	English Translation
Часть	621.21	A part
Вызывать	284.21	To cause / evoke
Действие	344.46	action / act
Оставаться	758.95	To remain / to stay
Развитие	379.93	Development
Народ	401.13	People / nation
Существовать	309.85	To exist
Власть	399.16	Power / authority
Уровень	382.11	Level
Считать	492.04	To consider
Земля	485.25	Earth / soil

FINDINGS AND INTERPRETATIONS

- The overall mean frequency decreases as the level goes up
- The overall mean frequency after removing extremely high frequency words has a reversal effect by level: the overall frequency increases as the level goes up
- High mean frequency of Level 1 is driven by very frequent words (e.g., most of which are function words)
- Level 1 has the highest variability of words (large SD)
- Level 1 contains thematic (immediate environment) words, which are typically lower frequency

MULTI WORD EXPRESSIONS

LEVEL 1

Multi-word Expressions	English Translation	Raw Frequency
По адресу	At the address	6
Может быть	Maybe / Perhaps	3
А также	As well as / also	2
С удовольствием	With pleasure/gladly	2

LEVEL 2

Multi-word Expressions	English Translation	Raw Frequency
Не только	Not only	8
По словам	According to (someone's words)	8
В результате	As a result	7
В принципе	In principle / Basically	6
А также	As well / Also	4
По поводу	Regarding / About	4
В том числе	Including	4
Во время	During	4
Как бы	As if / Kind of	3

LEVEL 3

Multi-word Expressions	English Translation	Raw Frequency
Тем не менее	Nevertheless / However	9
В общем	In general / Overall	9
В первую очередь	First of all	7
К сожалению	Unfortunately	7
С другой стороны	On the other hand	6
Дело в том	The thing is	6
На самом деле	Actually / In fact	5
По мнению	According to (someone)	5
С одной стоорны	On the one hand	4

MULTIWORD EXPRESSION CLASSIFICATION

Level 1

- Location
- Modality
- **Additive**
- Manner

Level 2

- **Additive**
- *Causal*
- *Resultative*
- *Temporal*

Level 3

- **Additive**
- *Causal*
- *Resultative*
- *Temporal*
- Contrast
- Concession
- Discourse Structuring
- Emphatic

FINDINGS AND INTERPRETATIONS

- As the level goes up, the number of multiword expressions increases
- Multiword expressions are functionally different at different levels
- The use of multiword expressions aligns well with the prototypical text types at different levels

GRAMMATICAL DESCRIPTIONS

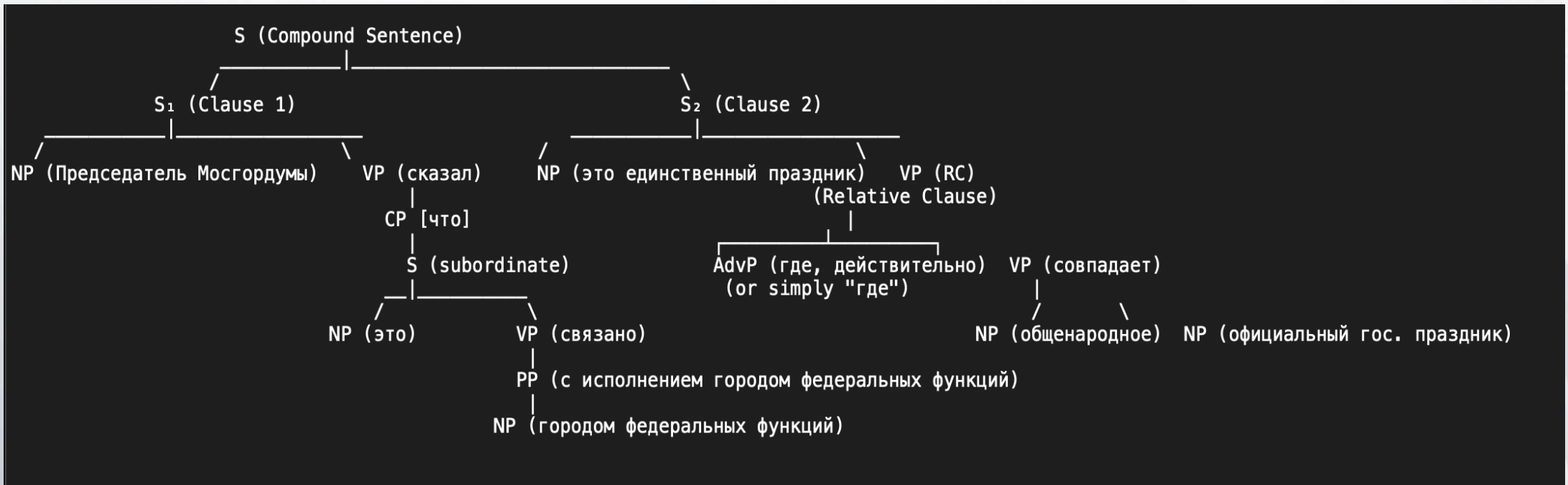
TEXT MEAN FREQUENCY METHODOLOGY

1. Texts were parsed for dependencies using Spacy
2. Tree depth was calculated for each text
3. Grand means of tree depth was calculated for each level
4. Words in texts were annotated with Parts of Speech (PoS) and other labels using Yandex Stemmer
5. Labels were counted per text and normalized per 100 words
6. Grand means of labels was calculated for each level

TREE DEPTH

$$\text{depth}(N) = 1 + \max(\text{depth}(L), \text{depth}(R))$$

Председатель Мосгордумы сказал, что это связано с исполнением городом федеральных функций, но это единственный праздник, где, действительно, совпадает общенародное настроение и официальный государственный праздник.



MEAN TREE DEPENDENCY DEPTH

Level	Mean	SD	Min	Max
Level 1	3.3	1	1.7	7
Level 2	3.6	.6	2.3	5.5
Level 3	3.9	.5	2.9	5.1

EMBEDDINGS

Председатель Мосгордумы сказал, что это связано с исполнением городом федеральных функций, но это единственный праздник, где, действительно, совпадает общенародное настроение и официальный государственный праздник.

The Chairman of the Moscow City Duma said that it is related to the city fulfilling its federal functions, but this is the only holiday where the general public's mood genuinely coincides with the official state celebration.

- **Clausal complexity (Subordination and coordination)**
 - Председатель Мосгордумы сказал ... но это единственный праздник
 - что это связано
 - где, действительно, совпадает
- **Phrasal complexity (prepositional phrases and noun phrases)**
 - с исполнением городом федеральных функций
 - общенародное настроение и официальный государственный праздник

DISTRIBUTION OF CASES PER 100 WORDS (YANDEX)

Level	Nom	Acc	Prep	Gen	Ins	Dat
Level 1	10	26	11	5	2	2
Level 2	9	22	13	6	3	3
Level 3	10	20	10	4	3	3

DISTRIBUTION OF VERB FORMS PER 100 WORDS (YANDEX)

Level	Perf Verbs	Imperf Verbs	Verbal Adv	Participl (Long)	Participl (Short)	Imper	Passive
Level 1	3	7	0.1	0.2	0.2	1.5	0.3
Level 2	7	7	0.2	0.9	0.7	0.2	1.2
Level 3	5	9	0.2	0.8	0.5	0.4	0.9

OTHER FEATURES PER 100 WORDS (YANDEX)

Level	Intro word	Particles	Geo name	First name	Last name
Level 1	0.7	3.3	2.3	0.7	0.3
Level 2	0.5	2.8	2.3	1.1	0.9
Level 3	1.1	5.3	1	0.6	0.7

Level	Imperatives	Intro words	Particles	Geo names	Participle (Short)	Participle (Long)
1	путешествуйт е одевайся экономь откройте купи			Белоруссия Казань Россия Москва Сочи Европа Италия		
2				Камчатка Анапа Ростов Подмосковье Брянск Новгород Ставрополь	обнаружен введен объявлено ужесточены освобожден госпитализирова н установлено доставлен	двигающийся рухнувшего построенные оборудованное находящимся расстрелявшего случившегося желающих
3		Фактически Действительно Собственно Вероятно Кстати Пожалуй Кажется	вот, ну, ли, не, только, просто, лишь, именно, разве, даже, бы		принято оформлены подписан корумпирован предназначены посвящена озвучено направлено	возобновляему имеющая касающемуся развивающийся действующие называемой связанные

Level	Imperatives	Intro words	Particles	Geo names	Participle (Short)	Participle (Long)
1	Travel Dress up Save Open up Buy			Belarus Kazan Russia Moscow Sochi Europe Italy		
2					Discovered Implemented Announced Tightened Released Hospitalized Established Delivered	Moving / in motion That collapsed Built Equipped Located Who opened fire That happened Those wishing
3		In fact Actually Essentially Probably By the way Perhaps It seems	Well, So, Whether, Not, Just, Only, Precisely, Really, Even		Adopted Formalized Signed Corrupted Intended Dedicated Announced Aimed / directed	Renewable That has Pertaining to Evolving Current / in effect Called Connected / related

FINDINGS AND INTERPRETATIONS

- As level goes up, so does the amount of structural/syntactic embedding
- Most used cases are Nominative, Accusative and Prepositional
- Accusative goes down as the level increases, but other cases remain stable at different levels
- Level 2 and 3 differ from Level 1 by the increased number of perfective verbs, verbal adverbs, short and long participles and passive constructions
- Level 1 is distinguished from Level 2 and 3 by a great number of imperatives

FINDINGS AND INTERPRETATIONS

- Level 3 is characterized by introductory words and particles compared to Level 1 and 2
- Levels 1 and 2 are characterized by a greater use of geographical names compared to Level 3
- Level 2 is characterized by a greater number of people's first and last names compared to Level 1 and 3

Listening Sub-Ability	Level 1 SLD	Our findings
Text Type	Understands connected phrases and sentences in simple speech. Has little to no ability to understand organizational devices.	<ul style="list-style-type: none"> • No organizational devices except for one type of additive
Lexical Control	<p>Understands some high frequency vocabulary.</p> <p>May show no understanding of major historical and cultural references.</p>	<ul style="list-style-type: none"> • Presence of extremely high-frequency vocabulary • Greater vocabulary variability • Thematic (immediate environment) low-frequency vocabulary • Multi-word expressions limited • Geographical names
Structural Control	Understands some basic structures.	<ul style="list-style-type: none"> • Imperatives • Nominative, Accusative and Prepositional • More imperfective than perfective verbs

Listening Sub-Ability	Level 2 SLD	Our findings
Text Type	Can understand speech that conveys sequencing of connected phrases. Has some understanding of frequently used rhetorical devices, such as repetition and exaggeration.	Additive, causal, resultative, temporal multi-word expressions
Lexical Control	Can readily understand high frequency vocabulary and sometimes understand references to major historical or cultural events.	<ul style="list-style-type: none"> • Less frequent vocabulary compared to Level 3 • Greater use of people first and last names
Structural Control	Generally understands information presented in almost all basic structures.	Greater use of participles (both short and long)

Listening Sub-Ability	Level 3 SLD	Our findings
Text Type	Able to understand cohesive discourse with clear relationships of ideas found in genres such as speeches, lectures, and briefings. Understands commonly used rhetorical devices such as simile and metaphor.	<ul style="list-style-type: none"> Slightly more frequent vocabulary compared to Level 2 Greater use and variety of multiword expressions (e.g., Contrast, Concession, Discourse Structuring, Emphatic)
Lexical Control	Readily understands a broad range of vocabulary, including mid frequency terms and technical vocabulary in the area of professional specialty. Understands widely used idiomatic expressions and cultural references.	<ul style="list-style-type: none"> Greater use of introductory words and particles
Structural Control	Understands a variety of structures accurately, but occasional misunderstandings occur in complex structures.	Heavier syntactical embedding (phrasal and clausal)

SUMMARY OF CONCLUSIONS

Level 1

- Extremely high-frequency vocabulary
- Greater vocabulary variability
- Thematic (immediate environment) low-frequency vocabulary
- Multi-word expressions limited
- Imperatives
- Geographical names

Level 2

- Less frequent vocabulary compared to level 3
- Greater use of people first and last names
- Greater use of participles (both short and long)

Level 3

- Slightly more frequent vocabulary compared to Level 2
- Greater use and variety of multiword expressions
- Greater use of introductory words and particles
- Heavier syntactical embedding

FUTURE DIRECTIONS

- Collect more texts for analysis
- Compare spoken vs written texts
- Add more features for analysis (e.g., phonological)
- Run a Multidimensional Analysis (MD) to uncover constellations of features
- Create pedagogical materials (based on the data)

THANK YOU! QUESTIONS!

CONTACT US AT: ALEKSEY.NOVIKOV@DLIFLC.EDU