

## A Quick Glance at ChatGPT and Languages other than English: Russian

Edna Andrews, Duke University

One of the topics that was not explored in our short commentary on ChatGPT published earlier this summer was interactions and outcomes in languages other than English (*Cognitive Neuroscience, Linguistic and Computer Science Perspectives of ChatGPT* (Glossos 18, Spring 2023 & IFLE v.9, issue 2, June 2023). The current short addendum, which is the first in a series of pieces devoted to world languages and ChatGPT, looks specifically at Russian language interactions in ChatGPT where the focus of the questions were about Russian language itself, as well as questions about the arts and popular culture. The same questions were asked at two different time points – April 8, 2023 and July 30, 2023. We compare the two sets of answers generated at these two time points and comment on relevant empirical and content-based outcomes. The overarching and most troubling outcome of these searches is the clear absence of normative Russian language usage in the answers, grammatical and lexical mistakes, and factual mistakes in both sets of longitudinal data. In general, the ChatGPT answers appear to be awkward and often inaccurate translations into Russian from another language, possibly English (see the Appendix for examples). It is important to keep in mind that existing large language models (LLMs) are primarily trained in English language data sets. This means that the current data sets for other languages of the world will be smaller, and this size difference will have implications for grammaticality and semantic outcomes in these smaller data sets.

### Language-oriented queries and questions about the arts and popular culture

The first question submitted to ChatGPT asks for a definition of verbal aspect in the *Russian literary language*. The term “literary” refers to the standard language as approved by the Russian Academy of Sciences Linguistics Institute [Институт языкознания Российской академии наук].

The explanations given in April and July differ in length and organization, but both are surprisingly incorrect and do not coalesce with any official descriptions by the Russian academy for describing the category of verbal aspect. Below are some examples:

- (1) Both versions claim that the perfective aspect requires additional suffixes in order to convert the imperfective form to perfective, specifically –l, -i, -e. In the first article, the claim is the addition of not only –l, but –la, -lo, -li. This is a clear confusion of **desinence** and **suffix**, as well as the verbal categories of **tense** and **aspect**.
- (2) The second version adds additional errors, including (a) calling the –l a suffix (it is a desinence/ending, NOT a suffix), (b) claiming that verbal aspect requires agreement with person, number and gender and (c) misrepresenting semantic outcomes. In verbal conjugations, the present tense form (which can only be imperfective) agrees in person and number, while the past tense form (which can be both imperfective and perfective) agrees in number and gender. In the final instance, while aspectual differences in verbs can be very different semantically, the examples given by ChatGPT are not (cf. видеть/увидеть is given as an example).

The July version has many more **grammatical** errors than the April version, and includes multiple examples where the prepositional/locative case is used when the accusative is required. Furthermore, instead of the word причастие, which as a grammatical term is used in the singular number, the term is used in the plural and in combination with the word “verb”.

A second question about the grammatical gender of the Russian words папа (“papa”) and кофе (“coffee”) evokes astounding responses from ChatGPT. Specifically, ChatGPT conflates grammatical and semantic gender, and also included answers that are not actually words in Russian at all (cf. баебонькание). Both version of ChatGPT completely miss the major point of the question, which requires an explanation about the grammatical importance of paradigmatic declensions, and fails to note that папа is a second declension masculine noun (it declines like a 2<sup>nd</sup> declension feminine noun, but has masculine syntagmatic agreement).

In the case of кофе, the answers from both versions are remarkable in their inaccuracy. The most glaring contradiction is that both version claim different genders – the first version gives only neuter, and the second only masculine. The first version states that “in everyday speech we often say ‘masculine coffee’, but that is probably a set

expression, not a grammatical rule.” This is absolutely incorrect. Finally, in the first version we correct the answer in saying that coffee is masculine, not neuter. ChatGPT answers immediately that we are right and it is wrong and apologizes for the error. However, as it tries to explain why it made the mistake, it compounds the error by claiming that кофе is indeed masculine because it ends in the vowel –e-. In fact, the word кофе is often mistaken as being neuter precisely because it ends in the vowel –e-. (Note that colloquial Russian recently recognized the use of neuter agreement with the indeclinable noun кофе.)

What is perhaps most interesting is that in the follow-up repeat of this question in July, 2023, ChatGPT emphatically picks “masculine” and provides an explanation based on a false etymology. The most respected *Etymological Dictionary of Russian Language* is the four-volume dictionary by Max Vasmer (reprinted 1986:355). Vasmer clearly notes the textual evidence of the form кофей from 1724 and notes that this word comes into Russian from English or Dutch. ChatGPT falsely claims it came directly from Arabic into Russian, which is incorrect. Even though ChatGPT lists the Arabic script of the word, it only provides one of the two variants of the transcription of the word.

Before moving on to a new question, ChatGPT also provides further food for thought in its musings about the word кофе in Russian. It falsely claims that кофе in Russian does not have a specific gender and can be used with masculine and feminine adjectives and pronouns in the April version, and provides additional nonfactual explanations in incorrect Russian about why кофе is masculine when the evidence provided is why it would be perceived as neuter. In the July version ChatGPT has a different explanation and states incorrectly that кофе is masculine precisely because of its historical origins. There are case errors in the July version (use of locative instead of the correct accusative case form for в этот период), as well as an ungrammatical sentence with both grammatical and lexical errors.

Another question was posed about a famous Russian actor of stage, film and TV – Maksim Averin. The question was simple: Who is Maksim Averin? The answers in both versions are mainly incorrect. The two versions give two different birthdays, and both are incorrect. Averin was at Satirikon Theatre in Moscow from 1997-2015 and later at Teatr Satiry (Theatre of Satire) from 2018 until the present. ChatGPT mentioned several

Moscow theatres, but not the two where Averin was part of the troupe. In terms of his films and TV roles, not a single title listed in either response is a part of Averin's filmography. In short, ChatGPT was 100% incorrect in both the April and July responses. This is quite surprising, but reminiscent of the nonexistent sources produced in response to queries made in our recent publication (Andrews et al. 2023, IFLE & Glossos). [For a correct listing of Maksim Averin's work in theatre, film and television, see [https://www.kino-teatr.ru/kino/acter/m/ros/31/works/.](https://www.kino-teatr.ru/kino/acter/m/ros/31/works/)]

### **Outcomes and future questions**

If we compare the two sets of responses from ChatGPT to the same questions posed in April and July, 2023, we find that while the actual answers are in many ways divergent, they are similar in repeating the same incorrect information on the one hand, and also making grammatical, morphological and lexical errors in Russian itself. If we focus specifically on grammatical case errors in the questions discussed above, the July version has many more errors than found in the April version. For readers interested in reliable sources on the Russian language, including grammar, morphology, lexicon and Slavic languages in general, please refer to the short bibliography included below. Further research is required to understand what the source of these errors is and how to correct them in future versions of ChatGPT.

## Bibliography on LLMs and ChatGPT

Meyer JG, Urbanowicz RJ, Martin PCN, O'Connor K, Li R, Peng PC, Bright TJ, Tatonetti N, Won KJ, Gonzalez-Hernandez G, Moore JH. ChatGPT and large language models in academia: opportunities and challenges. *BioData Min.* 2023 Jul 13;16(1):20. doi: 10.1186/s13040-023-00339-9. PMID: 37443040; PMCID: PMC10339472.

## Bibliography of prominent research on Russian language and Slavic languages

Borkovskij, Viktor I. and Petr S. Kuznesov. Историческая грамматика русского языка. 1965, Nauka.

Bräuer, Herbert. Slavische Sprachwissenschaft. Vol. 1-3. Berlin, De Gruyter, 1961-1969.

Gamkrelidze, Thomas V. and Vjaceslav V. Ivanov. Indo-European and the Indo-Europeans. A Reconstruction and Historical Analysis of a Proto-Language and Proto-Culture. Part I: The Text. Part II: Bibliography, Indexes  
(Translated by: Nichols Johanna. Preface by: Roman Jakobson)

Ivić, Pavle. Die serbokroatischen Dialekte : ihre Struktur und Entwicklung. Mouton, 1958.

Jakobson, Roman. Selected Writings. Vol. I-IX.

Lunt, Horace G. Old Church Slavonic Grammar (7th Revised Edition). Berlin/New York: Mouton de Gruyter, 2001.

Meillet, Antoine. Le slave commun. 2nd edition. Champion, 1965.

Trubetzkoy, N. S. Избранные труды по филологии. Progress, 1987.

Vaillant, André. Grammaire comparée des langues slaves. Vol. 1-5

t. 1. Phonétique.

t. 2. Morphologie. Ptie. 1. Flexion nominale; Ptie 2. Flexion pronominale.

t. 3. Le verbe.

t. 4. La formation des noms.

t. 5. La syntaxe.

## Appendix: (April answers in black, July answers in blue)

I. Questions: April 8, 2023 EA

### **Объясните грамматические правила, определяющие вид глагола в русском литературном языке.**

В русском языке существует два вида глагола: совершенный и несовершенный. Выбор вида зависит от характера действия, которое описывает глагол, и от контекста предложения.

Совершенный вид глагола используется, чтобы выразить действие, которое было завершено в прошлом или будет завершено в будущем. В русском языке совершенный вид образуется добавлением к корню глагола специального суффикса (-л, -ла, -ло, -ли), который указывает на завершенность действия.

Несовершенный вид глагола используется, чтобы описать действие, которое происходит в настоящем или происходило в прошлом, но не было завершено. Несовершенный вид образуется добавлением к корню глагола суффикса (-а-, -е-, -и-), который указывает на незавершенность действия.

Выбор между совершенным и несовершенным видом зависит от контекста предложения и того, какое значение нужно выразить. Например, если нужно описать действие, которое завершилось в прошлом, используется совершенный вид, например: "Я прочитал книгу". Если нужно описать действие, которое происходит в настоящем, используется несовершенный вид, например: "Я читаю книгу".

Также есть глаголы, которые могут иметь как совершенный, так и несовершенный вид. Например, глагол "писать" может быть как "написать" (совершенный вид), так и "писать" (несовершенный вид). В таких случаях выбор вида зависит от того, какое значение нужно выразить в конкретном предложении.

30 July 2023

### **Объясните грамматические правила, определяющие вид глагола в русском литературном языке.**

ChatGPT

В русском литературном языке глаголы могут быть разделены на два вида: глаголы совершенного вида и глаголы несовершенного вида. Вид глагола определяет, как описывается действие в отношении его завершенности или продолжительности. Для объяснения грамматических правил, определяющих вид глагола, рассмотрим основные характеристики каждого вида.

#### 1. Глаголы совершенного вида:

- Описывают действие, которое рассматривается как завершенное, цельное, выполненное до конца.
- Используются, когда акцент делается на результате действия или на самом факте его завершения.
- В совершенном виде глагола уточняется время завершения действия, например: я купил, он написал, мы съели и т.д.

- В совершенном виде образуются прошедшее время и повелительное наклонение глагола.

## 2. Глаголы несовершенного вида:

- Описывают действие, которое рассматривается как незавершенное, неопределенное по времени или продолжающееся.
- Используются, когда акцент делается на самом процессе действия или на его продолжительности.
- В несовершенном виде глагола уточняется время действия без указания его завершения, например: я покупаю, он пишет, мы едим и т.д.
- В несовершенном виде образуются настоящее время, будущее время и причастия глагола.

Грамматические правила, определяющие вид глагола:

### 1. Правило образования:

- Совершенный вид образуется при помощи специального суффикса -л-, -и-, -е- для прошедшего времени единственного числа и -л- для повелительного наклонения.
- Несовершенный вид образуется без добавления дополнительных суффиксов.

### 2. Правило согласования:

- В русском языке глаголы согласуются с подлежащим по лицу, числу и роду.
- При смене вида глагола у некоторых глаголов изменяется корень, и это может повлиять на форму слова и его окончания.

### 3. Правило использования:

- Выбор вида глагола зависит от контекста и смысла высказывания.
- Как правило, совершенный вид используется для описания завершенных действий и акцента на результате, а несовершенный вид - для описания незавершенных действий и акцента на процессе.

### 4. Особенности некоторых глаголов:

- Некоторые глаголы могут употребляться только в одном из видов, например: быть - несовершенного вида (я был, он будет), стать - совершенного вида (он стал).
- У некоторых глаголов меняется значение в зависимости от вида, например: видеть (несовершенный) - увидеть (совершенный), слышать (несовершенный) - услышать (совершенный).

Важно понимать, что выбор вида глагола не всегда очевиден, и это требует практики и опыта с русским языком. Знание грамматических правил поможет правильно использовать глаголы и тем самым сделать высказывание более точным и понятным.

### Почему слово "кофе" на русском языке мужского рода?

Существительное "кофе" в русском языке является среднего рода. Несмотря на то, что в повседневной речи мы часто говорим "мужской кофе", это скорее устойчивое выражение, чем грамматическая правда.

В русском языке есть несколько существительных среднего рода, у которых прилагательные и местоимения могут иметь как формы мужского, так и женского рода. Это связано с особенностями грамматики русского языка.

Таким образом, слово "кофе" на русском языке не имеет строго определенного рода и может употребляться с прилагательными и местоимениями как мужского, так и женского рода в зависимости от контекста и предпочтений говорящего.

Но слово кофе мужского рода.

Я прошу прощения, вы правы. Существительное "кофе" на русском языке действительно мужского рода. Я допустил ошибку в своем предыдущем ответе.

Это связано с тем, что в русском языке существительные могут быть мужского, женского или среднего рода в зависимости от их окончаний и других грамматических признаков. В случае со словом "кофе", его мужской род обусловлен окончанием "-е" и другими грамматическими признаками.

Извините за путаницу и неверную информацию в моем предыдущем ответе.

### Почему слово "кофе" на русском языке мужского рода?

ChatGPT

Существительные в русском языке имеют один из трех родов: мужской, женский или средний. Род слова "кофе" в русском языке — мужской, и это связано с историческими особенностями его происхождения.

Слово "кофе" в русский язык пришло из арабского языка "قهوة" (qahwah), которое, в свою очередь, возникло из слова "قَهَّ" (qahha) — "он заметил" или "он обратил внимание". Первоначально это слово обозначало напиток, который приготавливали из обжаренных зерен кофейного дерева.

В русский язык слово "кофе" попало в XVII веке через турецкий или польский язык. Именно в этом периоде было сформировано и закрепилось у слова "кофе" мужской род. В русском языке слова, заимствованные из других языков, могут унаследовать род, который был у этого слова в исходном языке.

Однако стоит отметить, что в ряде других языков, таких как французский или немецкий, слово "кофе" имеет женский род ("le café" и "der Kaffee" соответственно). Род слова может меняться в зависимости от языка и его исторического развития.

Таким образом, слово "кофе" в русском языке имеет мужской род из-за особенностей его исторического происхождения и заимствования из других языков.

### Кто такой Максим Аверин?

Максим Аверин - это известный российский актер, родившийся 26 мая 1975 года в Москве. Он окончил Российский государственный институт кинематографии им. С.А. Герасимова (ВГИК) в 1998 году.

Аверин начал свою карьеру в кино и на телевидении в 1990-х годах. Он снимался в ряде фильмов, таких как "Золушка" (1997), "Дом с лилиями" (1999) и "Ёлки" (2010). Он также стал популярным благодаря своим ролям на телевидении, включая сериалы "Кадетство", "Следствие ведут знатоки" и "Ментовские войны".

В последние годы Максим Аверин также занят работой в театре. Он играл в спектаклях МХАТ, Театра на Малой Бронной, Театра имени Моссовета и других театров.

### Кто такой Максим Аверин? Как называется его самая известная роль?

ChatGPT



Максим Аверин — российский актер театра и кино, родился 26 марта 1975 года. Он стал известен благодаря своим ярким ролям на телевидении и в кино, а также работе в театре.

Самая известная роль Максима Аверина — роль Олега Краснова в популярном российском телесериале "Карпов". Сериал выходил на экраны с 2009 по 2013 год и рассказывал о сложных отношениях между членами семьи Карповых, которые занимались строительным бизнесом.

В течение своей карьеры Максим Аверин снялся во многих кинофильмах и сериалах, таких как "Самый лучший фильм", "Мамы", "Саша+Маша", "Девчата", "Счастливы вместе" и других. Он также активно работает в театре и исполняет разнообразные роли на сцене.

Максим Аверин — один из популярных актеров России, его талант и харизма привлекают внимание зрителей, и он имеет широкую популярность в культурной среде страны.