Peer-reviewed article

Variation in translation: Entropy for Norwegian verbs based on Russian translations of the novel *Sofies verden*

Atle Grønn

Department of Literature, Area Studies, and European Languages, University of Oslo

atle.gronn@ilos.uio.no

Abstract

This pilot study explores translation variability between Norwegian and Russian verbs using entropy-based metrics. By analysing multiple Russian translations from the RuN-IT corpus, we quantify the (un)predictability of translation choices and identify various tense–aspect–mood (TAM) patterns that are of interest for contrastive grammar studies of Germanic and Slavic verbs. The findings from this pilot study confirm some earlier observations from research using parallel corpora, while at the same time identifying certain patterns overlooked by traditional methods. Methodologically, the article proposes verb entropy as a simple, reproducible metric for translation variation that can complement traditional parallel-corpus work.

1. Background

A recurrent problem with the use of translations in foreign language teaching and linguistic research is the reliability of the translation.[1] The data are typically, at best, obtained through parallel corpora, such as the ones compiled at the University of Oslo (going back to the 1990s, see Johansson (2007)):

(1N) #1 Sofie Amundsen var på vei hjem fra skolen. (first sentence of the novel *Sofies Verden* from 1991 by Jostein Gaarder, Norwegian original, source Oslo Multilingual Corpus)

[1] I'm grateful to the editors and reviewers for constructive feedback and to Victoria Li Fossum for helping me with programming in Python.

CC BY 4.0

(1E) Sophie Amundsen was on her way home from school. (English translation, OMC)
(1G) Sofie Amundsen war auf dem Heimweg von der Schule. (German translation, OMC)

For the RuN-corpus (2008-2010), we built further on OMC and added the Russian translation of *Sofies verden*:

(1R) София Амуннсен возвращалась домой из школы. (Russian translation, RuN)

For the so-called "Interaktiv fremmedspråkportal" (IFSP 2024-), currently under construction at the University of Oslo, we have also added the Ukrainian translation:

(1U) Софія Амундсен поверталася після занять додому. (Ukrainian translation, IFSP)

The above translations are all done by professional translators, published in their respective countries. To investigate how many acceptable solutions actually exist for a given sentence, we constructed the RuN Interactive Translation database (RuN-IT). Norwegian students of Russian (L3/L4, CEFR level varying from A2 to B2) translate short fiction excerpts (including *Sofies verden*) into Russian via an online interface and can compare their own solutions with several alternative translations made by Russians fluent in Norwegian. For the relevant excerpt from *Sofies verden* we have 11 such non-professional translations by native speakers in addition to the professional one in (1R).

For the new project IFSP, which involves migrating RuN-IT and adding new languages such as Ukrainian, German, French, and Spanish, we have also included AI translations made by DeepL and ChatGPT-4. In this article, I will instead focus on the "old" L1 data obtained for Russian through RuN-IT. RuN-IT contains about 1000 original Norwegian sentences (≈ 2500 verbs) and can be viewed as a small multi-retranslation corpus in the sense of Cheesman et al. (2017), who advocate using multiple independent translations of the same text to quantify variation and translator style. In this pilot study, I focus on one excerpt from *Sofies verden*, specifically 163 Norwegian verbs automatically identified by the Stanza POS-tagger in Python.

The primary contribution is methodological: I explore how word translation entropy can be used to quantify variation in multiple human translations of the same literary source text and how this quantitative signal can reveal subtle tense–aspect–mood (TAM) patterns that matter for contrastive grammar and L3/L4 instruction. The secondary contribution is descriptive: based on 163 Norwegian verb forms, I document several non-trivial TAM patterns in Norwegian vs. Russian (with occasional reference to Ukrainian), such as the use of Russian perfective infinitives under habitual marking and the Russian imperfective past for Norwegian experiential/universal perfects.

My focus on verbs can be justified on several grounds. Schultzen (2021) already did a study of aspectual competition and variation in the RuN-IT database (excluding

translations by L3/L4 students of Russian). The textbook[2] on translation from Norwegian to Russian by Nesset and Makarova (2016) contains chapters dedicated to verbal aspect (chapter 2), motion verbs (chapter 3) in addition to sections on modals (6.9), all of which are relevant for this work. In this respect, I follow the mentioned authors and focus on verbs in translation not primarily from a lexical point of view, but as a point of departure for the study of tense–aspect–mood-features (so-called TAM-categories).

2. Entropy in translation

The present study is explicitly exploratory and should be regarded as a pilot study of verbal categories. I do not test pre-registered hypotheses; instead, I "fish for results" in the sense of scanning the RuN-IT data for systematic entropy patterns that warrant more targeted hypothesis-testing in future work.

One reason for focusing on verbs is that they conveniently allow for the measurement of entropy. For each Norwegian verb in the original text, I sampled twelve Russian translations of that verb. I assumed that if no Russian verb was used in the translation, a null form had been used. Three additional assumptions were made: conservation of the number of verbs from the original to the translation and conservation of the order of verbs in the original and the translation. Finally, I assumed that the twelve translators provide us with all possible translations of the Norwegian verb in the original. While these assumptions clearly yield incorrect results at times, I opted against manual cleaning and post-processing of the data. The last assumption—that twelve translators collectively approximate an exhaustive set of acceptable Russian translations for a given Norwegian verb—is clearly unrealistic. It is adopted here as a modelling shortcut to make entropy computable on a small, fixed sample, and all results should be interpreted with this limitation in mind.

In the next step, I calculated in Python the *word translation entropy* (Bangalore et al. (2016), viz. the information value of the distribution over Russian translations of each Norwegian verb. Finally, I sorted the Norwegian verbs according to entropy to determine which Norwegian verbs were least ambiguous (lowest entropy) and which were most ambiguous (highest entropy) in the specific context of translation of *Sofies verden* to Russian.

Entropy measures the (un)predictability of a set of outcomes. In this study, the metric reflects the diversity of Russian verb forms (including 'null_V') used for each Norwegian verb. For details on the standard formula[3] for word translation entropy that

[2] In the mentioned textbook the Norwegian source text is constructed, and the authors provide only one translation into Russian — one "solution".

[3] These ideas famously go back to the American mathematician Claude Shannon (1916–2001), the father of information theory.

was used in Python, see Bangalore et al. (2016:214), Wei (2023:47), and Wang et al. (2025):

$$H(s) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

Low entropy (close to 0) indicates that the Russian translations for a given Norwegian verb are consistent, uniform and predictable, with one or a few verb forms dominating. High entropy indicates that there is a wide variety of Russian verb forms used to translate the Norwegian verb, with no single form standing out. However, one cannot, strictly speaking, infer from high entropy that the individual translator was hesitant and experienced difficulties in their choice of translation. Nevertheless, the word translation entropy is often used as a predictor variable to describe the lexical translation choices that a specific translator must make. This step rests on the assumption that the participants share the same mental lexicon, so that the entropy extrapolates from the translations made by different native speakers to the potential performance of one translator (Wei, 2023:46). In principle, one could use resampling techniques to derive approximate confidence intervals for each entropy value. Given the small sample size (n = 12), I leave such inference-oriented extensions to future work and restrict myself here to descriptive entropy comparisons.

Let me sum up the methodological limitations and potential biases of the study.

First, the RuN-IT dataset comprises twelve native-speaker translations of a short excerpt. This is rich by corpus standards but small relative to the conceptual space of acceptable translations.

Second, because the counter operates on surface strings, morphological variants such as *прошла* vs. *прошли* are treated as distinct types. In some cases, this is linguistically motivated—for instance, plural *прошли* in my data set signals a comitative construction that is unavailable in Norwegian. In other cases, string-level distinctions may inflate entropy without reflecting meaningful variation in TAM features.

Finally, to keep the procedure reproducible and scalable, I opted against manual cleaning of the data, even though this occasionally yields “messy” counters. The downside is that entropy sometimes mixes lexical variation, morphological variation, and tagging artefacts. A natural next step is to compute entropy over *features* (e.g. aspect, tense, finiteness) rather than over surface forms, which would both reduce sparsity in the data set and target the TAM categories of theoretical interest.

3. Discussion of RuN-IT translation data

I will present the data chronologically, as they appear in the novel. The average entropy for 163 verbs in the Norwegian original text is 1.6. The number does not mean much in isolation, but we keep it in mind in the discussion of the data.

(1N') #1 Sofie Amundsen var på vei hjem fra skolen. (Entropy of the Norwegian verb form "var" based on (1R'): 1.33.)
(1R') Counter #1 ('шла': 6, 'возвращалась': 5, 'была': 1)

The verb entropy 1.33 for #1 is slightly below the average entropy 1.6 in my dataset. The translations are basically reduced to two alternatives. Both verbs are imperfective, but the coarse-grained counter does not take aspectual morphology into consideration. The third alternative in (1R'), which only occurs once, is a calque from the Norwegian original, a typical case of "translationese" ('была по пути домой').

The second sentence of the novel is interesting from a temporal–aspectual point of view:

(2N) #2 Det første stykket hadde hun gått sammen med Jorunn. (Entropy of the Norwegian composite verb form 'hadde gått' based on (2R'): 1.28.)
(2E) She had walked the first part of the way with Joanna. (OMC)
(2R) Первый отрезок пути она шла вместе с Йорунн. (RuN)
(2R') Counter #2 ('прошла': 7, 'шла': 4, 'прошли': 1)

The past perfect auxiliary 'hadde' in the Norwegian original is tagged as AUX and not VERB by the Python-library Stanza, and thus not counted in the statistics, which is good for my purposes here since Norwegian frequently uses composite verb forms with auxiliary verbs, where Russian only has one verb.

Note that the counter in Python makes a distinction between strings, hence the plural 'прошли' is counted separately from the singular 'прошла'. On a constructional view (cf. the Tromsø project the Russian Constructicon (Bast et. al. 2021)), this choice can be justified since the use of plural is indicative of a comitative subject ('Первую часть пути они прошли вместе с Йорун'—translation from RuN-IT) which is a construction characteristic of Russian but not possible in Norwegian.

In (2R') we observe aspectual variance (cf. the study of Janda & Reynolds (2019)): 8 uses of the perfective vs. 4 uses of the imperfective. As noted by Schultzen (2021:38), 'идти' and 'пройти' do not normally constitute a trivial aspectual pair, but here this seems to be the case due to the presence of the direct object, the incremental theme 'первый отрезок/первую часть пути'. In fact, the use of the imperfective in (2R) comes close to a general-factual imperfective, while the alternative choice of the perfective more directly renders the resultant state component of the past perfective in the Norwegian original (see Grønn & von Stechow (2020) for a discussion of such aspectual translation patterns between Slavic and Germanic).

(3N) #9 Ved det store matsenteret hadde de skilt lag. (entropy 1.33)

(3E) When they got to the supermarket they went their separate ways. (OMC)
(3R) Возле продуктового центра, с его множеством магазинов и лавочек, дороги подруг разошлись. (RuN)
(3R') Counter #9 ('разошлись': 6, 'расстались': 5, 'попрощались': 1)

The official, published Russian translation differs slightly from the others since the subject in (3R) is inanimate 'дороги подруг', while in the 11 others, the subject is animate 'они (подруги)'. One could imagine that an inanimate subject makes the description less dynamic and would trigger imperfective aspect. This is what we get in the official Ukrainian translation ('their paths diverged'):

(3U) Біля великого супермаркету їхні шляхи розходилися. (IFSP)

A few sentences later, these aspectual subtleties come to fore:

(4N) #18 Her begynte den dype skogen. (entropy 0.00) #19 Nå svingte hun inn i Kløverveien. (entropy 1.19) #20 Helt innerst gjorde den en brå sving. (entropy 2.45)
(4E) This was where the woods began. She turned the corner into Clover Close. At the end of the road there was a sharp bend, (OMC)
(4R) Ведь прямо за их садом начинался густой лес. София свернула на Клёвервейен. Ближе к концу улица делала резкий изгиб, (RuN)
(4R') Counter #18 ('начинался': 12); Counter #19 ('свернула': 8, 'повернула': 3, 'завернула': 1); Counter #20 ('делала': 5, 'уходила': 2, 'был': 1, 'заворачивал': 1, 'заканчивалась': 1, 'имела': 1, 'сворачивала': 1)

The aspectual pattern in (4) is extraordinarily neat. The first (#18) and last (#20) sentences of (4N) are stative, backgrounded descriptions of the surrounding nature, and all Russian translations use the imperfective. For #18 all 12 translators choose the same form, 'начинался', giving verb entropy 0. This is pedagogically useful: despite the entropy signalling a highly expected and unambiguous choice, the imperfective may still not be obvious for L3/L4 learners, especially given the following sentence #19, where all translators switch to a perfective verb form to encode a dynamic action from Sophie's perspective. In #20, the verbs are again in the imperfective since the description relates to the static environment. The high entropy in #20 is due only to lexical variation; the consistent use of the imperfective is the main pedagogical takeaway.

The three sentences in (4) thus convincingly illustrate an aspectual rule in Russian: Static descriptions in the imperfective vs. dynamic events in the perfective aspect.[4] The same aspectual rule applies also in Ukrainian:

(4U) Тут починався дрімучий ліс. Ось вона звернула на Клевервейєн. Вулиця закінчувалася несподіваним закрутом, (IFSP)

[4] Of course, this rule depends on what we understand by "static descriptions vs. dynamic events". More generally, Janda and Reynolds (2019) found that in at least 17% of verbs in running text, native speakers rate both imperfective and perfective verbs as acceptable in the same contexts.

The next example is a complex sentence with several verb forms and intriguing aspectual properties:

(5N) #29 Hun pleide å legge (entropy: 2.28) en tjukk bunke [post/reklame] på kjøkkenbordet #30 før hun gikk (entropy: 2.52) opp på rommet sitt #31 og begynte (entropy: 3.09) på leksene.
(5E) [There was usually a lot of junk mail], a pile to dump on the kitchen table before she went up to her room to start her homework. (OMC)
(5R) Обычно София оставляла толстую пачку корреспонденции на кухонном столе и, поднявшись к себе в комнату, принималась за уроки. (RuN)
(5R') Counter #29 ('оставляла': 5, 'клала': 2, 'положила': 2, 'имела': 1, 'привыкла': 1, 'ложила': 1); Counter #30 ('подняться': 4, 'поднималась': 3, 'поднявшись': 1, 'пойти': 1, 'отправиться': 1, 'оставлять': 1, 'класть': 1); Counter #31 ('сесть': 2, 'начать': 2, 'делать': 2, 'принималась': 1, 'приступала': 1, 'пойти': 1, 'подняться': 1, 'начинала': 1, 'приступить': 1)

Syntactically, there are some important differences between Norwegian/English and Russian in this case. Note that the Norwegian original uses the habitual–aspectual auxiliary 'pleide – used to', which is absent in the Russian translations. Instead, we typically get an imperfective past in Russian, as 'оставляла' in (5R). As shown in (Grønn 2011), perfective aspect is not completely ruled out under habitual markers (for instance under 'бывало' or 'по привычке'). In the words of Schultzen (2021:44): «The translators opting for the perfective aspect single out a specific occurrence of an habitual action. […] As such, 'по привычке она клала...' translates to 'habitually she put…', while 'по привычке она положила…' translates to 'out of habit she put…'.»

What is particularly striking in (5R') is the distribution of finite vs. non-finite (infinitive, gerunds) verb forms. The pattern that emerges from these data is that the perfective aspect is used despite the habitual context when the verb is non-finite (#30 'подняться': 4; 'поднявшись': 1; #31 'сесть': 2, 'начать': 2; 'приступить': 1), while the imperfective past is typically preferred for finite verb forms under habitual operators (#30 'поднималась': 3, #31 'принималась': 1, 'приступала': 1, 'начинала': 1).

The use of perfective aspect with non-finite verb forms in these contexts has to my knowledge not been properly described in the literature (but see Dickey (2000) for some relevant discussion). The more expected imperfective past is consistently used in the official Ukrainian translation:

(5U) Дівчинка завжди клала чималеньку купу пошти на кухонний стіл, а тоді піднімалася до себе в кімнату і бралася до уроків. (IFSP)

For perspicuity, I provide below two of the non-professional Russian translations with non-finite perfective verb forms in combination with the imperfective past:

(5R'') Обычно София оставляла толстую пачку корреспонденции на кухонном столе и, поднявшись к себе в комнату, принималась за уроки. (RuN-IT)
(5R''') По привычке она клала толстую кипу на кухонный стол перед тем, как подняться в свою комнату и сесть за уроки. (RuN-IT)

Another topic in verb semantics which is of interest in a cross-linguistic Slavic-Germanic setting, is the use of modals. Among the 163 verbs in the dataset, the modal ‘burde’ in (6N) gets the highest possible entropy:

(6N) #97 Er det allikevel noe som burde interessere alle? (entropy: 3.59)
(6E) Is there nothing that interests us all? (OMC)
(6G) Gibt es trotzdem etwas, das alle interessieren sollte? (OMC)
(6R) А нет ли какого-нибудь предмета, который бы заинтересовал всех? (RuN)
(6R’) Counter #97 ('бы': 1, 'касающееся': 1, 'смогло': 1, 'будет': 1, 'может': 1, 'интересующее': 1, 'могло': 1, 'было': 1, 'интересовало': 1, 'вызывать': 1, 'существовать': 1, 'null_V': 1)

For sentence #97 every single string of each translator is unique. This is unsurprising, since there are no clear one-to-one correspondences between modals in Germanic and Russian (cf. Yanovich (2013), ch. 5 on deontic modals), and ‘burde’ is one of several hard nuts to crack in translation of Norwegian modals (cf. Nesset & Makarova 2016:152). We also note that the Norwegian modal is ignored in the English translation (6E) but rendered in the German one in (6G).

For reference, I also give the Ukrainian translation which is rather similar to the original Russian one in (6R):

(6U) А чи є щось таке, що зацікавило б усіх? (IFSP)

Note that in my statistics (6R) and (6U) would be treated differently due to the order of ‘by’/‘b’ with respect to the finite verb. Furthermore, it is not obvious that the subjunctive particle ‘by’/‘b’ should be tagged as a VERB, but this choice by Stanza makes sense here with a Norwegian modal verb in the original. From a difficult modal, I turn to a simple indicative sentence:

(7N) #112 Hva er viktigst i livet? (entropy: 0.92)
(7E) What is the most important thing in life? (OMC)
(7R) Что важнее всего в жизни? (RuN)
(7R’) Counter #112 ('null_V': 8, 'является': 4)

The translator into Russian is effectively left with two options. The example shows that I need a null verb category in the counter. In the next example, the follow-up to the question in (7N), the verb entropy is even lower:

(8N) #113 Hvis vi spør (entropy: 0.82) en som lever på sultegrensen, er svaret mat.
(8E) If we ask someone living on the edge of starvation, the answer is food. (OMC)
(8G) Wenn wir jemanden in einem Land mit Hungersnot fragen, dann lautet die Antwort: Essen. (OMC)
(8R) Если мы спросим об этом у человека, живущего на грани голода, он ответит: еда. (RuN)
(8R’) Counter #113 ('спросим': 10, 'зададим': 1, 'Спроси': 1)

The very low entropy of ‘spør’ in #113 reflects the clear dominance of 'спросим'; the two other candidates are mere outliers. Interestingly, all 12 translators use a perfective

verb (future or imperative), whereas the Norwegian original has a present tense in the antecedent of a conditional. This is a non-trivial observation, and it confirms a more general difference between Germanic languages (Norwegian, English, German) and Slavic languages (Russian, Ukrainian) in the use of present vs. future in adjunct clauses (cf. Grønn & von Stechow 2011).

The Slavic pattern with a future in the adjunct is also attested in the Ukrainian translation:

(8U) Якщо ми запитаємо людину, яка помирає з голоду, вона відповість — хліб, (IFSP)

This pattern is once more confirmed in the *if*-adjunct in #116:

(9N) #116 Hvis vi retter (entropy: 2.05) det samme spørsmålet til en som fryser, er svaret varme.
(9E) If we ask someone dying of cold, the answer is warmth. (OMC)
(9G) Wenn wir dieselbe Frage an einen Frierenden stellen, dann ist die Antwort: Wärme. (OMC)
(9R) Если мы обратимся с тем же вопросом к замерзающему, он скажет: тепло. (RuN)
(9U) якщо ми звернемося з таким запитанням до того, хто замерзає, відповідь буде — тепло. (IFSP)
(9R') Counter #116 ('зададим': 5, 'обратимся': 3, 'адресуем': 2, 'Обратимся': 1, 'Спроси': 1)

The entropy is much higher for #116 than #113 because of the wide choice of lexical verbs in the counter. However, the important point for us is that all the verb forms are perfective (11 future; 1 imperative), while the Germanic versions are all in the present tense. The Russian verb 'адресуем' is biaspectual but must here be interpreted as perfective in analogy with the other forms in the dataset.

In the next example, the three first verbs all have high entropy and pose interesting problems for the translators:

(10N) #145 Den som interesserer seg for slike spørsmål (entropy: 3.02), #146 er opptatt av noe (entropy: 3.42) #147 som mennesker har diskutert (entropy: 3.02) så lenge vi har levd på denne planeten.
(10E) People who ask such questions are taking part in a debate that has gone on as long as man has lived on this planet. (OMC)
(10G) Wer sich für solche Fragen interessiert, beschäftigt sich mit etwas, das die Menschen schon fast so lange diskutieren, wie wir auf diesem Planeten leben. (OMC)
(10R) Люди обсуждают подобные проблемы столько тысячелетий, сколько населяют нашу планету. (RuN)
(10U) І хто цікавиться цим питанням, приєднується до мільйонів тих, хто добирав способу до його розв'язання, відколи людство існує на цій планеті. (IFSP)
(10R') Counter #145 ('интересуется': 3, 'обсуждают': 2, 'Интересующегося': 1, 'стали': 1, 'бьются': 1, 'ставит': 1, 'занимают': 1, 'задаётся': 1, 'задаёт': 1); Counter #146 ('занимается': 2, 'населяют': 1, 'занимает': 1, 'интересовать': 1,

'озабочен': 1, 'существуют': 1, 'интересуется': 1, 'заинтересован': 1, 'занят': 1, 'обеспокоен': 1, 'null_V': 1); Counter #147 ('null_V': 3, 'обсуждают': 2, 'обсуждали': 1, 'увлекались': 1, 'тревожит': 1, 'спорили': 1, 'Волнуют': 1, 'обсуждается': 1, 'обсуждало': 1)

The official Russian translation in (10R) is a simplification compared to many of the other 11 Russian translations in (10R') and compared to (10U). What is of particular interest in (10N) is the switch from the present tense in #145 and #146 to the universal perfect ('har diskutert' 'a debate that has gone on') in #147. In (Grønn & von Stechow 2020), we discuss parallel corpus translations of the universal perfect from English to other European languages (cf. also van der Klis et. al (2022), who use multilingual narrative corpora and quantitative methods to study the competition between past and perfect in Germanic and Romance languages). The Germanic languages are not uniform here, as we see in German (10G) with an inclusive present tense ('diskutieren'). Ukrainian (10U) actually uses an experiential imperfective past 'добирав' in combination with an inclusive present 'існує'. Several of the Russian translations use present tense forms such as 'обсуждают' or 'обсуждается', in accordance with the findings of (Grønn & von Stechow 2020) and a rule proposed in (Nesset & Makarova 2016), to which I return in (12N) below.

In the discussion of (5) above, we encountered 'unexpected' perfective infinitives in Russian. The next example illustrates nicely how difficult the aspectual choice can be in the infinitive:

(11N) #152 Den beste måten å nærme seg filosofien på (entropy: 2.75), #153 er å stille noen filosofiske spørsmål (entropy: 2.45): Hvordan er verden skapt?
(11E) The best way of approaching philosophy is to ask a few philosophical questions: How was the world created? (OMC)
(11R) Входить в философию лучше всего постановкой философских вопросов. Как устроен мир? (RuN)
(11U) Найкращий спосіб наблизитися до розуміння філософії — це поставити кілька філософських питань: Як створено світ? (IFSP)
(11R') Counter #152 ('приблизиться': 4, 'быть': 2, 'Входить': 1, 'подойти': 1, 'сблизиться': 1, 'постичь': 1, 'познакомиться': 1, 'задать': 1); Counter #153 ('задать': 5, 'поставить': 2, 'устроен': 1, 'ответить': 1, 'ставить': 1, 'создан': 1, 'задавать': 1)

The entropies for both Norwegian infinitive verbs are high, and the results are rather unclear, also when it comes to aspect. The official Russian translation uses an imperfective infinitive 'Входить', while (11U) makes use of two perfective infinitives. Overall, the counter in (11R') shows a preference for perfective infinitives also in the Russian non-professional translations.

(12N) #163 Spørsmål som disse har vært stilt av mennesker til alle tider. (entropy: 2.45)
(12E) People have been asking these questions throughout the ages. (OMC)
(12G) Solche Fragen haben die Menschen zu allen Zeiten gestellt. (OMC)

(12R) Такие вопросы задавались людьми каждой эпохи. (RuN)
(12R’) Counter #163 ('задавались': 5, 'задавали': 2, 'ставили': 1, 'интересовались': 1, 'занимали': 1, 'волновали': 1, 'интересовали': 1)

The final sentence of the excerpt contains a present perfect in the Norwegian original. Apparently, ‘har vært stilt’ in (12N) is a universal perfect, which is used to denote situations that obtain both in the past and at the utterance time. For this reading of the perfect, Nesset & Makarova (2016:55) formulate the so-called “Perfektumregelen”, a pedagogical rule which says that you should always translate a universal perfect in Norwegian into present tense in Russian. This rule is widely used in the teaching of Russian grammar at Norwegian universities, but the rule is seriously challenged by the data in (12R’): Not a single translator uses the present tense!

Perhaps one could argue that (12N) has a flair of both the universal perfect and the so-called experiential perfect. In a similar case above (10N/10R’), we encountered both present tense and imperfective past in the translations. For some reason, all the Russian translations in (12R’) have a past imperfective, which is the preferred choice for the translation of experiential perfects, as noted in both Nesset & Makarova (2016) and Grønn & von Stechow (2020).

4. Conclusion

By analysing multiple Russian translations from a quantitative perspective with verb entropy, I have identified some new patterns and questioned some established wisdom in the TAM domain of contrastive Germanic–Slavic studies. We can sum up the discussion of the selected sentences above from lowest to highest verb entropy as in figure 1.

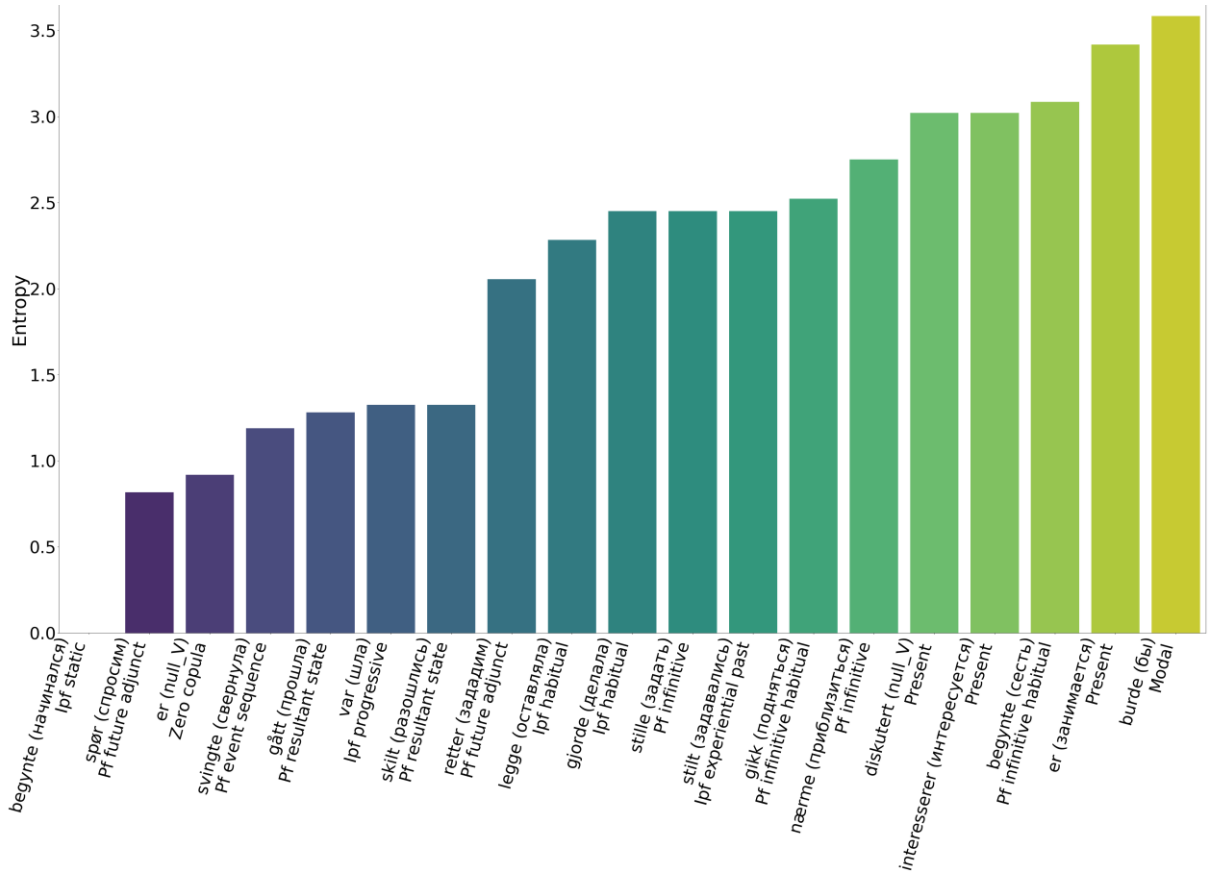


Figure 1. Entropy of Norwegian verbs given Russian translations.

In particular, the RuN-IT data highlight two phenomena: (i) the use of perfective aspect with non-finite verb forms under habitual marking (#30–#31, #152–#153), and (ii) the use of the imperfective past in contexts where the Norwegian original has a universal perfect (#163), challenging the so-called *Perfektumregelen* (Nesset & Makarova 2016).

Low entropy means that there is a clear preference for a specific Russian translation. However, low entropy does not automatically imply that the choice is trivial for L3/L4 learners. For instance, choosing 'спросим' in #113 or 'начинался' in #18 is presumably not obvious for Norwegian learners, despite the low verb entropy of the Norwegian originals.

The empirical base for this study is deliberately narrow: a single excerpt from one novel by one author. As such, the entropy values and TAM patterns documented here are not intended to generalize directly to other genres or registers. Rather, they demonstrate the potential of multi-translation data; systematic comparison across additional texts is left for future research.

This pilot study was explicitly «fishing for results». In the discussion I have focused on phenomena that have received some attention in previous studies using parallel corpora (e.g., RuN; Parasol (von Waldenfels 2012) etc.), but I did not test specific hypotheses, and the statistics are based on only moderately accurate tagging by the

Stanza library in Python. Future work should compute entropy over features rather than over surface forms and should test specific TAM-hypotheses on larger multi-retranslation corpora.

The multi-translation data in RuN-IT was sampled before the AI-revolution, which adds to its value since pure human data can be hard to obtain these days outside of strictly experimental environments. Human translations are expensive. Thus, in machine learning research, one rarely uses more than four human translations of the same text (Victoria Li Fossum, p.c.).

Jostein Gaarder's novel *Sofies Verden* is a case where authentic fiction can serve pedagogical purposes in translation practice (at, say, B1-level). The 163 verbs analysed here are usually presented to students in chunks of four. In the spirit of Ilya Frank's method for foreign language teaching (Mammadova & Belyaeva 2020) and *the Russian Constructicon* (Bast et al. 2021), there is added value in using authentic texts even in classroom settings.

For L3/L4 learners, the key pedagogical payoff of verb entropy is that it highlights zones of systematic variation (e.g. modals, perfects, non-finite forms) where it is didactically useful to present learners with multiple acceptable Russian solutions instead of a single "correct" translation. In addition, there are some non-trivial cases where we can actually point to a unique "correct" translation, notably concerning aspectual choice with stative descriptions (Ipf) vs. dynamic events (Pf).

References

Bangalore, S., Behrens, B., Carl, M., Ghankot, M., Heilmann, A., Nitzke, J., Schaeffer, M., and Sturm, A. 2016. Syntactic variance and priming effects in translation In *New Directions in Empirical Translation Process Research*, 211–238. New York: Springer.

Bast, R., Endresen, A., Janda, L.A., Lund, M., Lyashevskaya, O., McDonald, J., Mordashova, D., Nesset, T., Rakhilina, E., Tyers, F.M., Zhukova, V. 2021. *The Russian Constructicon. An electronic database of the Russian grammatical constructions*. Available at https://constructicon.github.io/russian/.

Cheesman, T., Flanagan, K., Thiel, S., Rybicki, J., Laramee, R.S., Hope, J., Roos, A. 2017. ulti-Retranslation corpora: Visibility, variation, value, and virtue. *Digital Scholarship in the Humanities* 32(4). 739–760. https://doi.org/10.1093/llc/fqw027.

Dickey, S.M. 2000. *Parameters of Slavic Aspect. A Cognitive Approach*. Stanford: CSLI Publications.

Grønn, A. 2011. 'Byvalo' and 'Used to' as Verbal Quantifiers. *Slovo* 52. 63–80.

Grønn, A. & von Stechow, A. 2011. Future vs. present in Russian and English adjunct clauses. *Scando-Slavica* 57(2). 245–267. https://doi.org/10.1080/00806765.2011.631783.

Grønn, A. & von Stechow, A. 2020. The Perfect. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann & Thomas Zimmermann (eds.), *The Wiley Blackwell Companion to Semantics*. John Wiley & Sons. https://doi.org/10.1002/9781118788516.sem046.

Janda, L.A. and Reynolds, R.J. 2019. Construal vs. redundancy: Russian aspect in context. *Cognitive Linguistics* 30(3). 467–497. https://doi.org/10.1515/cog-2017-0084

Johansson, S. 2007. *Seeing through multilingual corpora*: *On the use of corpora in contrastive studies.* Amsterdam: John Benjamins.

van der Klis, M., le Bruyn, B., and de Swart, H. 2022. A Multilingual Corpus Study of the Competition between Past and Perfect in Narrative Discourse. *Journal of Linguistics* 58. 423–57.

Mammadova, N. & Belyaeva, V. 2020. Advantages and disadvantages of learning foreign languages according to Ilya Frank's reading method based on the example of Russian students, In *INTED2020 Proceedings*, 8144–8148.

Nesset, T. & Makarova, A. 2016. *Ikke bare-bare. En liten bok om å oversette til russisk.* Oslo: Novus.

Schultzen, V.M. 2021. *Aspectual competition and variance in translation*. MS thesis. University of Oslo. https://hdl.handle.net/10852/87913

Wang, Z., Cheung, A.K.F., Xu, H. et al. 2025. Assessing lexical and syntactic simplification in translated English with entropy analysis. *Humanit Soc Sci Commun* 12. Article 1213. https://doi.org/10.1057/s41599-025-05562-9

Von Waldenfels, R. 2012. ParaSol: Introduction to a Slavic Parallel Corpus. *Prace Filologiczne* LXIII. 293–302.

Wei, Y. 2023. *Ambiguity and Entropy in the Process of Translation and Post-editing*. Diss. Dublin City University.

Yanovich, I. 2013. *Four pieces for modality, context and usage*. Diss. Massachusetts Institute of Technology.