

# Commercially Minor Languages and Localization

Maria Ivanova Todorova  
Sofia University  
mivanovat@uni-sofia.bg

## Abstract

This paper offers a perspective of languages with a less significant volume of digital usership as minor in the context of globalization and localization. With this premise, the risks this status poses to the quality of localized texts, the substantiality of genre conventions, the public image of professional translators, and the users' linguistic competence in these languages is explored. Furthermore, the common lack of established or clear conventions in the localization of digital products into commercially minor languages (and in the digital product genres) is highlighted as one of the factors amplifying these risks. These perspectives are contextualized with the Bulgarian language with examples of errors encountered in Bulgarian digital content localized from English and more specifically – errors and problems related to gender neutrality and register.

**Keywords:** minor languages, localization quality, linguistic competence, English-Bulgarian translation

## 1 Introduction

In localization, a major language is considered a language with high global distribution and commercial use, and a minor one – the opposite. In this context, not only all minority languages become minor, but also the understanding of “minority” in the globalized context should be re-examined. As early as 20 years ago, Cronin makes one crucial claim that is denotative for linguistic inequality in an era of technology and globalization, by declaring that “the hegemony of English in the fastest-growing areas of technological development means that all other

languages become in this context, minority languages.” (2003: 146). Speaking about this extended idea of “minority”, Cronin lists several negative consequences of “the absence of a critical self-reflective activity in the translation enterprise”: the vulnerability of minority languages to interference in majority-to-minority translation, the alteration of minority languages in translation, and the predominant focus on majority languages in Translation Theory (2003: 149). This paper applies Cronin's logic by extension to Bulgarian, a globally and commercially minor language, and, considering the all-pervasiveness of digital content in modern everyday life, explores the potential issues such a lesser status can lead to in the context of digital product localization.

## 2 Localizing digital content in Bulgarian

In the globalized present, every digital-market player strives to bring their product to as many users as possible – a crucial step in this process is localization, the adaptation of said product to the target market's locale<sup>1</sup>. When international enterprises enter the small Bulgarian market, they are first faced with one dilemma – to localize or not to localize. The inevitability of this question lies in “the quantitative prism of return on investment”, through which business executives perceive localization (Dunne, 2006: 1).

One can easily imagine that with fewer than 5 million users who access the Internet on a monthly basis (Miniwatts Marketing Group, 2023) Bulgaria is not such an enticing venture. Nevertheless, out of 22 identified English-source websites in the 100 most visited by Bulgarian users in August 2021, 12 have been localized into Bulgarian. What is more, out of the 7 originally English ones with over 1 million unique visits by Bulgarian users in the same

---

<sup>1</sup> ‘not just a language, but usually a particular variety of a language, plus local conventions regarding currency, date and hour settings, presentation of numbers, right through to such things as symbolic color coding’ (Pym, 2004: 2).

period, only 1 has not been localized (SimilarWeb, 2021).

What does this mean for the Bulgarian-speaking usership? Firstly, a great share of Bulgarian users find it easy to navigate online content and services in English, with localization remaining a lower priority than the significance of the service offered through this content. Secondly, despite the existence of such a big fraction being able to navigate in English, more than half of the companies which have their services based on English still choose to localize their content in Bulgarian. Among those are the giants Microsoft & Google each of which own dozens of digital service software products, as well as most of the “big players” with significant market reach, such as Booking, or eSky. They do so regardless of the uncertain or potentially small benefit from this investment. The reason behind this could be an effort towards achieving a universal and inclusive market reach. The lesser commercial status, however, combined with the complexity and multimodality of localization as a process, puts Bulgarian and other commercially minor languages in a volatile position when it comes to the investment in and, respectively, quality control on localized content.

What is more, when discussing the pragmatic quality of translation solutions in localization, Jimenez-Crespo refers to the conventionality in genres and translation adequacy in finding corresponding conventions in source and target languages (Jimenez-Crespo, 2009: 68). Even if some conventions in the digital product genres exist in Bulgarian and other minor languages, however, they have been heavily influenced by localization practices with “the majority of software and web applications [were] being developed [in English] in the United States” during the dawn of localization (Esselink, 2000: 4). In other words, most underlying genre conventions have originally appeared in English and then have been established in other languages through finding functionally equivalent solutions to already existing paradigms in English originals. When there is no full correspondence between these paradigms and those of the target language, translators in localization projects end up in the critical position of the ones to determine and establish the good practices.

As Postolea puts it, “especially when their target language is lesser known and terminologically

standardised, specialised translators, just like journalists or authors, are creators of language too [...] From this perspective, specialised translators have responsibilities towards the target language too and they should be wary of the risk of introducing loose translations and poor style into the target context.” (2016: 63) This responsibility, however, rarely corresponds to proportionate resources (both economic and informational) in localization to a minor language, which inevitably leads to diffuse accountability and, respectively, questionable quality. It is also a prerequisite for companies to rely more heavily on automation and AI-powered machine translation (MT) – an approach which boosts inclusivity, but is still far from ideal for smaller and low-resource languages (Pym, Ayvazyan, Prioleau, 2022: 13).

### **3 Translation errors in digital products localized into Bulgarian**

Translation errors in localized websites, mobile applications, or even operation systems are certainly not uncommon and we often laugh about them in professional circles and even blame them on machine translation. MT, however, is rarely exclusively used in the localization context. In order to analyze and systematize the types of errors found in products localized from English into Bulgarian, I collected 150 mistakes found in 42 such products by browsing and using my mobile devices in Bulgarian for a little over a year, excluding mistakes from products that were highly likely to have used non-post-edited MT exclusively or errors that cannot be linked to factors specific to the context of localization. Some of the errors I encountered could still be attributed to MT, especially when it comes to short ambiguous or polysemous user-interface text units. The ones included in the corpus, however, were found in products whose companies have shared information about their localization processes publicly, often showing how they approach and value quality assurance – among those are various Microsoft and Google products (including Android), Revolut, Glovo, AirBnB, and others. Regardless of whether and to what extent MT was used in the processes that led to these mistakes, their existence and nature bring up concerns about the average level of quality in localization from English into Bulgarian and the potential consequences thereof.

The errors encountered range from problems related to text types or communicative situations typical of the digital product genres (37 %) to problems linked to the technological macrostructure behind the texts showing in digital products (41 %) and problems resulting from various levels of ambiguity of the succinct source texts (22 %) (Appendix A). Due to volume constraints, for the purposes of this paper, I will focus on the former type – genre-related errors, and more specifically, those related to gender neutrality and register (Appendix B), while some examples of the other categories can be found in Appendix C.

The mistakes associated with those two aspects occur mainly due to the asymmetry between English and Bulgarian in the categories of gender and number, and more particularly in lexemes referring to the user in the second person. Judging by the mistakes encountered, there does not seem to be a universal approach in translating digital product text types with such lexemes, which leads to inconsistency (Image 1) and grammatically incorrect solutions which hundreds of thousands of users witness on a daily basis. Considering the minor status of Bulgarian in localization, motivation to tackle these issues would unlikely come from the top (product owners) and Postolea’s statement that translators become “creators of language” rings all the more true.

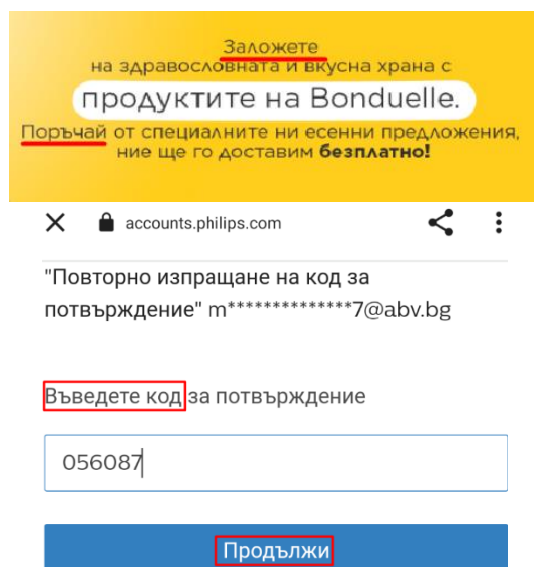


Image 1: Register inconsistency in Glovo and Philips Home

What happens though when there does not seem to be a consensus among stakeholders? Regardless of how common such translation inconsistencies

and mistakes can be in localized products, it is unlikely they would lead to any drastic linguistic changes on their own. In some of the linguistics aspects they are associated with, however, faltering of the norms (or uncertainty where one lacks) is observed on a greater scale.

The difficulties of adopting gender-neutral language in Bulgarian, for instance, often spark debates in professional translators’ groups. It might seem that we are far from codifying a solution, but translators have to find one on a regular basis – be it when localizing a product addressing an unknown individual user or a book or movie with non-binary characters. In localization, the lack of established strategies to apply in such cases mainly leads to ungrammatical solutions and inconsistency. If one solution were to prevail, however, it can slowly become an “unspoken rule” in localization practice and in digital product genres, if not in more universal use. In such context, as owners of the target language, translators involved in localization projects become decision-makers with the two most obvious choices of either disregarding the original neutrality or adapting the respective paradigm in Bulgarian to match it. As a result of this reality, the target texts in the minor language often end up as either too conservative and sometimes outright incorrect (Appendix B: Lines 21–24) or the target language simply gives way to the paradigm of the English source, using plural forms when referring to an individual user (Appendix B: Lines 10–20). Undoubtedly, these are not the only solutions when a variety of translation shifts can be adopted to bypass the asymmetry – e.g. instead of “Ще бъдете пренасочени” [You will be redirected(pl.)], a structural shift can be adopted to avoid the use of participle which will have to be marked for gender in the singular: “Ще Ви пренасочим” [We will redirect you(sg.)]. Such strategies, however, might not be universally applicable and often introduce additional limitations in the localization context that already implies numerous constraints on its own as seen in the second error category in Appendix A. Nevertheless, they can be applied in most of the context of the observed errors but have not, which lends the industry and academia at least one actionable solution – to include such strategies in translator training for the purposes of localization.

While the matter of linguistic gender-neutrality is yet to be solved in Bulgarian, the polite form of

address has already been codified. Despite the various adjustments over the years, we now have clear rules in regard to the use of capitalized pronouns and plural forms. And even though the past faltering of this form has weakened its use in present day and might have led to some of the related errors found in localized products, there are consistent patterns found that are likely due to something more than a disregard of the rules. In fact, Bulgarian localizers often oscillate between the informal singular and formal addresses, with the majority of products resorting to a “hybrid” non-codified form, characterized by the use of lowercase pronouns and plural adjectives and past participles when referring to an individual user. (Todorova, 2023: 122–123) What is more, judging by Google and Microsoft’s localized texts, this hybrid version is consistent with the conglomerates’ style guides and requirements. Unfortunately, Google does not provide open access to its documentation, while Microsoft only mentions the matter in passing, not singling it out, perhaps due to its lack of official codification (Todorova, 2023: 124).

#### 4 User reception

The reception of such translation errors would depend on the linguistic competence of the user. If they are a proficient speaker, such errors could only erode the image of professional translators. On the one hand, if the user is aware that a human is indeed behind such an error, it raises the question of how well professional translators are trained and how much these services are valued by investors in the face of digital product owners. On the other hand, if the user assumes the mistake results from the use of MT, the notion that professional translation services are not valued and sought after is reinforced. And in the cases where the user’s linguistic competence is already compromised, encountering such mistakes in the everyday use of their phone, for instance, could only deepen their lack of understanding or solidify previously adapted incorrect paradigms.

Additionally, if users commonly encounter such unfortunate localization instances, the question stands of how many people would actually choose to navigate digital content in their native tongue instead of in English when this is possible. In a country such as Bulgaria with a digitally minor language and an above-average rate of English proficiency (Education First, 2023: 4), proficient

English speakers would likely prefer to use digital products in English. As a result, even though they might have a better than average command of Bulgarian as well, their digital experience will be shaped by English modalities.

What is more, in a study on digital users’ reception, Taanonen concludes that “the [user] group with competent English skills is relatively more critical towards translated content and its quality than the group with weaker English-language skills.” (2014: 94) Losing even a part of this proficient and critical towards translation quality group, decreases the chance of motivation for improvement coming from the users. Additionally, by turning to English, users of higher proficiency decrease the market value of Bulgarian further and, respectively, the incentive for product owners to invest in improving the quality of localization.

#### 5 Final remarks

In the economic enterprise which localization essentially is, the quality of target texts in a language of a commercially minor status such as Bulgarian is often of lesser investment concern. This leaves translators at the vulnerable position of decision-makers within a low-resourced specialized context. Changes to this reality in the localization process, however, can mainly be expected to stem from the stakeholders considered owners of the minor language – the translators and translator trainers. A step in this direction could be the forming of a better understanding of translation problems in localization from English into Bulgarian, which would lead to the establishing of practical solutions and their integration in translation training.

On a grander scale, with the rise of AI-powered MT, which outperforms any human linguist in terms of speed and cost-efficiency, all the more focus should be put on finding ways to educate users and ensure their linguistic skills are strong enough to sift digital content through and identify quality localized content. Additionally, efficient processes should be found to guarantee that stakeholders in Bulgarian localization, such as translators and translation agencies, receive proper training, are equipped with sufficient resources, and enjoy professional prestige that would motivate higher responsibility and performance.

## References

- Michael Cronin. 2003. *Translation and Globalization*. London and New York: Routledge.
- Anthony Pym. 2004. *The Moving Text.: localization, translation, and distribution*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Keiran J. Dunne. 2006. Introduction: A Copernican revolution. In *Perspectives on Localization*, pages 1–11. Amsterdam/Philadelphia: John Benjamins Publishing.
- Miniwatts Marketing Group. 2023. *Data for the 27 Member States of the European Union*. Accessed on 14 April 2024. <https://www.internetworldstats.com/europa.htm#bg>
- Sorina Postolea. 2016. Translating in a Specialised Context: Challenges and Risks. In *Buletinul Institutului Politehnic din Iași* (Bulletin of the Polytechnic Institute of Iasi), LXII(LXVI), f.1, pages 51–66.
- SimilarWeb. 2021. *Industry Leaders Data*. Accessed on 22 September 2021. <https://pro.similarweb.com/>
- Miguel A. Jimenez-Crespo. 2009. The evaluation of pragmatic and functionalist aspects in localization: towards a holistic approach to Quality Assurance. In *The Journal of Internationalization and Localization*, 1(1): 60–93. Amsterdam/Philadelphia: John Benjamins Publishing.
- Bert Esselink. 2000. *A Practical Guide to Localization*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Anthony Pym, Nune Ayvazyan, Jonathan Prioleau. 2022. Can raw machine translation ensure inclusion? The case of public-health information in Catalonia. *1<sup>st</sup> International Conference Right to Languages: Linguistic Policies and Translation and Interpreting in Public Services and Institutions*.
- Maria Todorova. 2023. Forms of Address in Website and Software Localization From English into Bulgarian. In: *Studia Philologica Universitatis Tarnovensis (series annua)*, 42. Veliko Tarnovo: Veliko Tarnovo University Publishing House, 119–126.
- Education First. 2023. *EF English Proficiency Index. A Ranking of 113 Countries and Regions by English Skills*. Accessed on 14 April 2024. <https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2023/ef-epi-2023-english.pdf>
- Mari Taanonen. 2014. *Targeting Language Markets through Web Localization: A Case Study of a Digital Cloud Service*. Master's Thesis. Espoo: Aalto University

**Appendix A Problem categories and respective error types found in products localized into Bulgarian**

CATEGORY	SUBCATEGORY	TYPE
GENRE-RELATED	FORM OF ADDRESS	INCONSISTENCY
		LACK OF PRONOUN CAPITALISATION IN THE POLITE ADDRESS
		ADJECTIVE OR PARTICIPLE PLURALISATION
	USER GENDER	ADJ. OR PARTICIPLE MASCULINISATION
		ADJ. OR PARTICIPLE NEUTERISATION
		INCONSISTENCY
	CALL-TO-ACTION NAVIGATION TEXTS	AMBIGUITY
		INCONSISTENT TERMINOLOGY
	TERMS & NEOLOGISMS	INCONGRUITY AND/OR AMBIGUITY OF TERMS
MACROSTRUCTURAL LIMITATIONS	VARIABLES IN THE SOURCE CODE	WORD ORDER INTERFERENCE
		VARIABLE-SPELLING OF BULGARIAN PREPOSITIONS
		ASYMMETRY IN GENDER
		ASYMMETRY IN NUMBER
	CHARACTER LIMITATIONS	NON-VISIBLE USER-INTERFACE (UI) TEXT DISRUPTING THE PRODUCT'S FUNCTIONALITY
		VISIBLE UI TEXT DISRUPTING THE PRODUCT'S DESIGN
		UNCLEAR ABBREVIATIONS
	LOCALE CONVENTIONS	DURATION
		CURRENCY
		CAPITALIZATION
INSUFFICIENT CONTEXT	& SOURCE AMBIGUITY	POLYSEMY
		FUNCTIONAL POLYSEMY
		HETEROSEMY
		SYNCRETISM
		SYNONYMY
	& TARGET PARADIGMATIC VARIATION	ERRORS DUE TO ASYMMETRY IN GENDER
		ERRORS DUE TO ASYMMETRY IN NUMBER
		ERRORS DUE TO ASYMMETRY IN DEFINITENESS
		FUNCTIONAL POLYSEMY IN TARGET TEXT
		PUNCTUATION INTERFERENCE

Appendix B Some of the errors related to register and gender neutrality

No.	ENGLISH SOURCE	BULGARIAN TARGET	PRODUCT	TYPE
1.	Thanks for reviewing your stay, Maria!	Благодарим <b>ви</b> , че оценихте престоя си, Maria!	Booking	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
2.	Log in as Maria... Not <b>you</b> ?	Влизане като Maria... Не сте <b>вие</b> ?	Facebook	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
3.	Manage <b>your</b> Google Account	Управление на профила <b>ви</b> в Google	Google Chrome	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
4.	Maria your photo book draft is about to expire	Maria, черновата на албума <b>ви</b> е на път да изтече	Google Photos	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
5.	This app is an internal test version that has been shared with <b>you</b> . If <b>you</b> install it, <b>your</b> email address will be shared with the developer.	Това приложение е вътрешна версия, която е споделена с <b>вас</b> . Ако го <b>инсталирате</b> , имейл адресът <b>ви</b> ще бъде споделен с програмиста.	Google Play	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
6.	{username} liked <b>your</b> post	{username} хареса публикацията <b>ви</b>	Instagram	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
7.	Autofill <b>your</b> card details at checkout	Автоматично попълване на данните на картата <b>ви</b> за плащане	Revolut	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
8.	An activation email was sent to <b>your</b> email address.	Имейл за активация беше изпратен до <b>вашия</b> имейл адрес.	Xiaomi Home	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
9.	Every night when you are on Wi-Fi, we'll automatically download <b>your</b> favourite music...	Всяка вечер при наличие на връзка с Wi-Fi ще изтегляме любимата <b>ви</b> музика...	YouTube Music	LACK OF CAPITALISATION IN SINGULAR POLITE ADDRESS
10.	You will be <b>forwarded</b> back to the home screen	Ще бъдете <b>пренасочени</b> към началния екран.	BigBlueButton	ADJECTIVE OR PARTICIPLE PLURALISATION IN SINGULAR POLITE ADDRESS
11.	Are you <b>sure</b> you want to follow this link?	<b>Сигурни</b> ли сте, че искате да последвате тази връзка?	Facebook	ADJECTIVE OR PARTICIPLE PLURALISATION
12.	Are you <b>sure</b> you want to log out?	<b>Сигурни</b> ли сте, че искате да излезете?	Facebook	ADJECTIVE OR PARTICIPLE PLURALISATION
13.	{user} commented on a photo that you're <b>tagged</b> in	{user} коментира снимка, в която <b>сте отбелязани</b>	Facebook	ADJECTIVE OR PARTICIPLE PLURALISATION
14.	Have an account already?	Вече <b>сте регистрирани</b> ?	Glovo	ADJECTIVE OR PARTICIPLE PLURALISATION
15.	Are you <b>hungry</b> ?	<b>Гладни</b> ли сте?	Glovo	ADJECTIVE OR PARTICIPLE PLURALISATION
16.	<b>Unstoppable!</b>	<b>Неудържими</b> сте!	Google Fit	ADJECTIVE OR PARTICIPLE PLURALISATION

17.	{user}, are you ready to print your photo book?	Мария, <b>готови</b> ли сте да отпечатате фотоалбума си?	Google Photos	ADJECTIVE OR PARTICIPLE PLURALISATION
18.	{user} commented on a post you are tagged in	{user} коментира публикация, в която <b>сте отбелязани</b>	Instagram	ADJECTIVE OR PARTICIPLE PLURALISATION
19.	Do you know and trust {user}? If you are <b>unsure</b> , don't pay them...	Ако не сте <b>сигурни</b> , не му плащайте...	Revolut	ADJECTIVE OR PARTICIPLE PLURALISATION
20.	–	Здравейте, Мария, Имаме страхотни новини: официално сте <b>проверени</b> . Това означава, че вече можете да използвате своята карта на Revolut...	Revolut	ADJECTIVE OR PARTICIPLE PLURALISATION
21.	We've contacted you because you opted-in to receive newsletters from Takeaway.com	Свързахме се с теб, защото си се <b>съгласил</b> да получаваш новини от Takeaway.com	Takeaway	ADJECTIVE OR PARTICIPLE MASCULINISATION
22.	Maria, are you ready to...	Мария, готов ли си да се впуснеш...	Takeaway	ADJECTIVE OR PARTICIPLE MASCULINISATION
23.	Followed by {user}	Последвано от {user}	Instagram	ADJECTIVE OR PARTICIPLE NEUTERISATION
24.	Do you know and trust {user}? If you are unsure, don't pay <b>them</b> ...	Познавате ли и имате ли доверие на <b>{female user's name}</b> ? Ако не сте сигурни, не му плащайте...	Revolut	ADJECTIVE OR PARTICIPLE MASCULINISATION

Appendix C Some examples of errors form other categories

ENGLISH SOURCE	BULGARIAN TARGET	PRODUCT	CONTEXT
Weekly on {day of the week}	Ежеседмично в вторник	Google Calendar	Incorrect spelling of the preposition <i>o</i> in the settings of an event due to the variability of the succeeding noun.
1B+ Downloads	1 млрд.+ <b>Изтегляния</b>	Google Play	Use of English capital case in Bulgarian.
{number} key moments in this video	7 ключови моменти в този видеоклип	Google + YouTube	Use of incorrect plural form of a masculine noun due to asymmetry between English and Bulgarian in the category of number in combination with a variable.
{number} timer set	1 таймер е зададен	Android	English word order in info text potentially due to hardcoded macrostructure.
{number} days left	17 дни остават	Fundraising platform	English word order in info text potentially due to hardcoded macrostructure.
Order no.	Номер на поръчка	Obuvki.bg	Omission of the definite article in Bulgarian
Hi {user},	Здравей Мария,	Takeaway	Adopting the English source's punctuation in the context of a variable element.



---

**SPECIAL SESSION ON  
WORDNETS, FRAMENETS AND ONTOLOGIES**

---