



ARTICLE

Expanded View of Non-nativeness in Cyber Risk Perception

William Smeal, L. Jean Camp, and Mihai Păunescu *

Email: wsmeal@iu.edu

(First published online September 7, 2026)

Abstract

With the wider integration of the Internet outside of Western countries, questions have arisen about the role of language in social engineering and trust. Early phishing attacks targeted English-speaking victims largely in the United States. Even now, some of the most available, widely used, and well-documented indicators of cyber risks stem from exemplars of American nineties phishing scams. Perhaps the most indicative of the user-facing characteristics are related to writing style and usage, particularly those associated with non-native speakers. The Internet has also had a profound effect on non-Western literary practices as the Internet infrastructure (e.g., domain names, programming languages) is grounded in the Latin alphabet. For most of the world population this has necessitated the adoption of non-native writing styles. For any communities that do not use Latin script, non-native writing practices are an integral part of simply using the Internet. Yet risk perception in these linguistic communities has not been widely explored. In this work, we gauge the effects varying degrees of non-native style and usage have amongst eastern-Europeans and their intuitions about cyber risks. The results of this study can contribute to a more holistic analysis of risk perception, one that includes diverse linguistic and cultural perceptions of risk and trustworthiness in digital communication.

Keywords: phishing, cyber risk perception, non-nativeness, digraphia, Latinization, Cyrillic

1. Introduction

Despite the notoriety of email phishing attacks such as the Nigerian Prince scam, this type of internet criminality remains a high-volume activity. Furthermore, the increasingly ubiquitous nature of digital internet communication across the world facilitates the high-volume nature of these attacks, as malicious actors can now target nearly any population around the globe. As human users constitute the final line of defense against the success of phishing attacks, studies addressing phishing resilience must also take into account certain aspects of cultural perceptions and resilience.

Analyzing cultural perceptions is particularly applicable to research on phishing email scams because style and usage play instrumental roles in both attackers' attempts to steal sensitive information, as well as end users' attempts to avoid falling victim to online scams. A majority of phishing email resilience work focuses on English language attacks where writing style and usage is among the most influential gauges of potential risks [DDMP21]. This work seeks to expand upon the perception of non-nativeness and its influence in determining cyber risks and untrustworthiness. The *raison d'être* of this work stems from two distinguishing factors: 1) the status of English

as a global lingua franca, which significantly increases contact between native and non-native writers and 2) the infrastructure of internet protocols and computing languages, which necessitate the adoption of non-native writing practices amongst native speakers in certain linguistic communities.

This work seeks to explore these two particularities and their effect on cyber risk perception. Given the disproportionate amount of English as a second language speakers, recipients of phishing emails in less-common languages will be less accustomed to non-native style and usage errors. This study also offers an opportunity to investigate a new frontier in human mindsets and online risks: the advent of digraphic writing and its perceived legitimacy. To increase the understanding of human decision-making heuristics in relation to common cyber risks, this study aims to address the following research questions:

RQ1: How do native speakers of non-Latin languages interpret style and usage errors within the context of cybersecurity and online risks?

RQ2: Do instances of non-nativeness in online texts evoke a significant sense of risk or illegitimacy at similar rates across different linguistic communities?

RQ3: Do instances of digraphia suggest that internet users see changes to traditional writing as an objective product of modern technology or an inherently improper element indicative of insecurity and illegitimacy?

II. Related Work

Writing style and usage is an instrumental element of determining legitimacy in digital written communication and its importance is cited repeatedly in email phishing resiliency studies [Fur07, PBP⁺16]. More specifically, the general characteristics that differentiate legitimate and phished emails are consistency and personalization of the message, perceived legitimacy of links and sender, and the presence of spelling or grammatical irregularities [Sch12]. The latter two characteristics can be interrelated and will be the focus of the present study.

In work detailing the linguistic characteristics of email scams, Schaffer [Sch12] highlights the role of perceiving non-nativeness in determining the legitimacy of email communication. It is important to point out, however, that even native speakers commit certain grammatical and mechanical errors. For example, some of the most common errors committed by native English speakers include improper use of tense and conjugations, improper use or omission of the apostrophe, and comma errors [Sch12]. While some native readers may or may not be able to identify these common mistakes, a perception of non-nativeness develops when combinations of grammatical, mechanical, and or semantic misuses of the target language are repeatedly compounded. The perception of non-nativeness accentuates a lack of familiarity, which in turn can call into question the trustworthiness of the sender and the legitimacy of their written word.

The English language has consistently remained informationally dominant since the formation of the internet [Sun10]. Compiling data from 189 countries, Wolk [Wol04] correlates English language proficiency with the rate of Internet users per Internet Service Providers (ISP). Developed countries with high English proficiency yielded 188,000 users per ISP while developed countries with low English proficiency yielded 713,000 users per ISP. The divide in developing countries was even starker: developing countries with high English proficiency yielded 1,829,000 users per ISP while developing countries with low English proficiency yielded 10,233,000 users per ISP. Citing these figures, Wolk [Wol04] describes what has become known as “the digital divide”, an assertion that English language dominance has created a de facto linguistic protocol for the growth of the Internet. As a result, the role of perceived non-nativeness is well-documented in studies pertaining to phishing identification and resilience in English language communication. However, as addressed by Guo and Yao-Chung [GF24], there exists a dearth of research in multi-lingual phishing emails.

Even in the early 2000s native English speakers accounted for just about a quarter of Internet users. Although the English language remains informationally dominant online, a significant minority of internet users are genuine native speakers. English has more L2 (English as a second language) speakers than any other language in the world [BVK⁺15]. As a result, contact between L1 (native language) and L2 English speakers (or writers) is exceptionally high in comparisons to native/non-native interactions in other tongues. The implications of this disparity in terms of cybercrime such as phishing remains relatively unexplored. As native English speakers have accumulated a higher degree of exposure to writing containing non-native style and usage, the present study aims to test the receptiveness to non-native style and usage amongst linguistic communities with lesser degrees of interaction between native and non-native writers.

For the purposes of this experiment, the linguistic communities selected are native speakers of the Bulgarian and Russian languages. This choice in linguistic community offers two advantages. First, it allows for a tiered comparison of the perceptions of non-nativeness in cyber risk perception. As English represents the linguistic community in which interactions between native and non-natives is highest, Russian language allows for the analysis of a more niche linguistic community that also retains a degree of sustained exposure between native and non-natives. Although to a lesser degree than the anglophone community, the Russian language boasts a considerable amount of L2 speakers, particularly in Central Asian countries such as Tajikistan, as well as Caucasian countries such as Armenia and Georgia. The Bulgarian linguistic community represents an even more niche group within which contact with non-native speakers is very uncommon. A secondary advantage stems from the literary history of these two languages, the Cyrillic script, which allows us to analyze the relationship between digraphia and cyber risk perception.

Among the foundation of Internet infrastructure protocols exist a series of linguistic characteristics, which have the capability to restrict certain communicative interactions to exclusively Latin characters. Standard Mail Transfer Protocol (SMTP), for example, restricts address names to only Latin characters a-z, regardless of case, and Arabic numerals 0-9 [Ria07]. Additionally, the informational and communicative dominance of the English language has entrenched English (and its Latin alphabet) as the language in which the majority of digital information and universally-used software is written. The underlying technological framework utilized to produce and access the ever-increasing amounts of digitalized text have, nonetheless, resulted in adaptations to the traditional writing practices of respective regional scripts. Accessing certain domains and creating an email address are two examples of common internet actions for which internet users writing in scripts such as Armenian, Cyrillic, or Devanagari (Hindi) must make a manual switch to Latin script in order to complete. Instances in which users repeatedly switch keyboards may prompt a sort of “linguistic economy” whereas to avoid constant interruptions, the use of Latin is continued even when writing in a language traditionally written in another script. Under such circumstances digital writing has given rise to a notable increase in digraphia – the use of two scripts to write a single language. Other notable drivers of digraphia include the use of “partial bilingualisms” – the adoption of a foreign word (typically English) into another language where the script of the foreign word is retained. Similarly, the practice of “dual adaptation” in marketing and advertising – the retention of a brand name without transcription – is particularly common for Western brands so that the image associated with a particular brand name remains consistent and recognizable anywhere in the world [Kir18]. Although writing in two scripts is inherently non-native in languages such as Bulgarian, Georgian, or Korean, contemporary text production technology facilitates, and sometimes requires this practice.

III. Experimental Procedure and Design

This study aims to evaluate the relationship between writing style and usage and phishing resilience among communities beyond the Anglosphere, particularly in the southern Balkans and

Russian-speaking countries of eastern Europe. The participants in this study are native speakers of one or more Slavic, Cyrillic languages from Bulgaria, Russia, and Ukraine. The experiment begins with a survey consisting of questions regarding computer knowledge and skills, as well as computer security knowledge. The questionnaire is followed by the main experiment, in which participants are presented with images of written emails, and participants indicate whether or not they consider the respective email to be legitimate or a potential scam. After completing the principal experiment, participants answer a brief demographic survey. The questions constituting the initial survey are available in the appendix, accompanied by a translation in English.

Previous research indicates that people make decisions about legitimacy based on a variety of cues in digitalized writing [PBP⁺16]. The emails shown to participants in the present study include a combination of the elements outlined in the following list:

1. Links & URLs
2. Logos
3. Professional appearance
4. Personalization
5. Spelling and Grammar
6. Security Indicators
7. Sender Identity
8. Importance & Urgency

The study was conducted with a total of 52 participants aged 18 and over. All participants are native speakers of Bulgarian, Russian, and or Ukrainian. Table 1 shows the distribution of participant demographics:

Table 1. Distribution of participant demographics by region and country

Country	Participants	Region
Bulgaria	22	Southern Cyrillic Orthographic Zone
Russia	25	Eastern Cyrillic Orthographic Zone
Ukraine	5	Eastern Cyrillic Orthographic Zone

Participants were shown fifteen emails in total, all of which contained some form of URL or link to open another page. The purpose of including this detail is to simulate a tactic often used by cyber criminals in which a malicious link is embedded in the body of an email in order to steal personal information or install malware on a user's device. For control purposes, ten emails were classified as legitimate and therefore presented a professional appearance, included personalized greetings, displayed logos of the sending company or organization, contained no spelling nor grammatical errors, and refrained from language creating a sense of urgency. All but one of the legitimate emails were written in the Bulgarian and Russian languages using Cyrillic script; one was digraphic.

Three emails were classified as illegitimate, or phishing emails, and were presented in a plain text format, included general, impersonalized greetings, displayed no logos or imagery, contained some spelling and/or grammatical errors, and exercised an urgent tone to express a sense of time sensitivity. Common grammatical and spelling errors were included such as the omittance of commas before the conjunctions *че* and *что* (that), as well as any declined forms of the pronouns *който* and *который* (that, which). Additionally, the text of the phishing emails contained spelling errors such as incorrect cases of capitalization regarding any declined forms of the pronouns *ви* and *вы*, (you) as well as the possessive pronoun *ваш* (your).

A key part of this research addresses the question of digraphia and the increase of its prevalence in Cyrillic cyberspace. As implied by research question three, there is no single circumstance

that determines an instance of Cyrillic-Latin digraphia. As a result, recipients of such messages may contribute the change in script to a variety of factors. For example, perhaps the device from which the message was sent is not the sender's personal device, but rather a public computer with no Cyrillic keyboard installed. Similarly, certain older mobile devices are equipped with solely Latin characters, rendering digraphic emails reminiscent of early SMS messaging where the Latinization of Cyrillic characters was an obligatory practice.

It is important to note that Latinizing Cyrillic languages such as Bulgarian, Russian, and Ukrainian cannot, by definition, be done correctly. There is no official transliteration between Cyrillic and Latin. As a result, individuals create their own mappings resulting in inconsistencies. However, all native speakers of Cyrillic languages will have been confronted with Latinization for one reason or another by virtue of the Anglocentric nature of the internet and, by effect, the technology utilized to produce digital communication. In order to shed light on this area, this study attempts to gauge participants' perceptions of digraphia in the context of safety, trustworthiness, and legitimacy online. To this end, two "Latinized test" emails are included in the principal experiment. A Latinized test email displays all the characteristics of legitimate emails as described above, but the text of the email is entirely transcribed in Latin characters. These emails should be classified as legitimate. The extent to which they are identified as phishing is the test of RQ3.

IV. Results

The analysis of the results of this study addresses the research questions listed in Section I. The responses to the initial survey questions serve as a series independent variables utilized to gauge the target populations' computer and security expertise, risk assessment behavior, and phishing resilience. These variables are utilized to provide more in-depth insights into the perception of online security, legitimacy, and trustworthiness in relation to the phenomenon of Latinization. Statistical tests are conducted in order to identify any significant differences in risk perception among participants from the southern COZ and the eastern COZ (Cyrillic Orthographic Zone).

A. Phishing and Security Certificate Knowledge

The first two survey questions address participants' understanding of cybercrime and security, particularly knowledge of phishing and security certificates (questions 1 and 2 in the appendix). Below each question is listed a series of multiple-choice questions. The responses were recorded on a scale from 0 to 5. For example, question 1 asks "What is phishing?" and participants were instructed to choose all possible correct answers: 'Pretending to be someone or a company to steal user information', 'Making a fake website that looks legitimate to steal user information', 'Sending spam emails', 'Defrauding someone online', 'Impersonating someone or a company to steal user information', 'Tracking your internet habits to send advertisements', and 'I don't know'.

In order to account for partially correct responses, the final score was calculated using the following formula: $(\#CorrectOptionsSelected + 1) / (\#IncorrectOptionsSelected + 1)$. This same method was previously utilized in the development of a similar security expertise measure [MXR23]. Mehrabi et al. [MXR23] built on this expertise model in creating an experimental harness for phishing evaluations; (however, there is no support for Cyrillic in that framework). A similar approach was used by Kelley and Betenthal [KB16] in cognitive science, and by Das in her evaluation of the effect of age on phishing resilience (Das) and in cross-national comparisons [GC21] (Camp et al.).

The distribution shows that phishing and security certificate knowledge are generally low across both populations, but with some regional differences. The results of the one-way ANOVA test show that levels of phishing knowledge are significantly different between the southern and eastern COZ populations (p-value 0.01). Here, the eastern COZ group displays greater knowledge of phishing. In terms of security certificate knowledge, the results of the one-way ANOVA test show that levels of knowledge are not significantly different between the two groups (p-value 0.33).

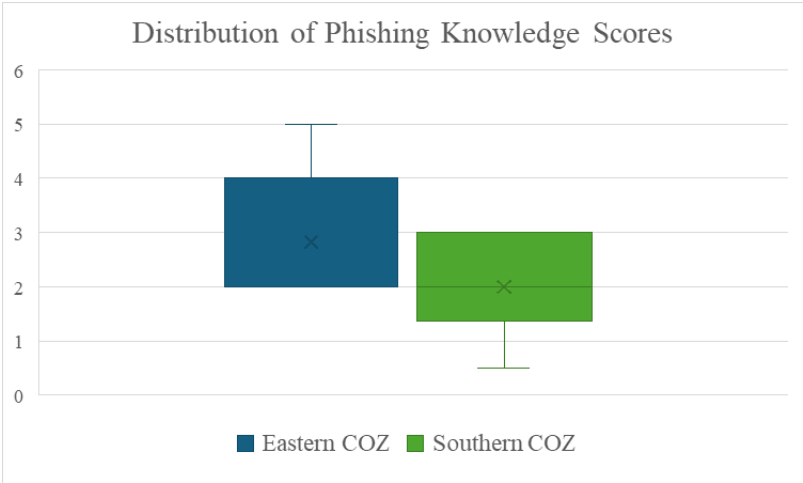


Figure 1. Distribution of phishing knowledge scores.

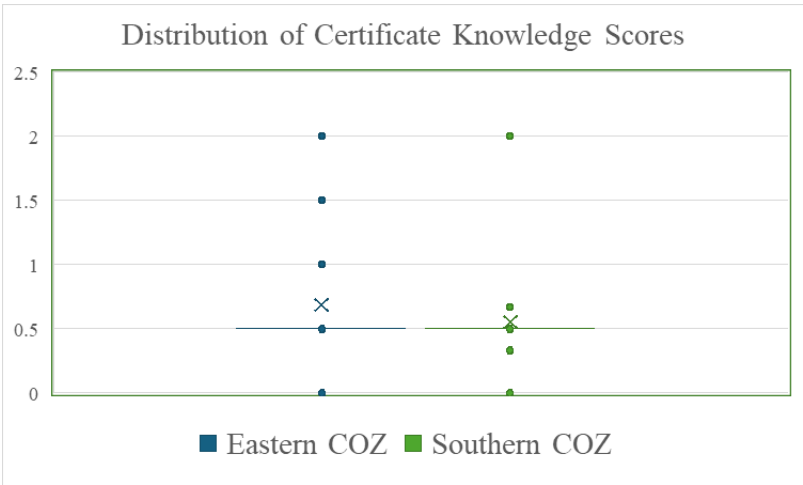


Figure 2. Distribution of certificate knowledge scores.

B. General Cybersecurity Expertise

The following three questions (questions three, four, and five in the appendix), test participants’ knowledge of general computer security [GC21]. Participants were awarded one point for each correctly answered security knowledge question. Each question presents only one correct answer. The results of this segment of the survey showed that general security knowledge is relatively low among both groups of participants in this experiment. This finding is particularly true among the southern COZ, where general security knowledge is significantly lower than that measured in the eastern COZ (p-value 0.9).

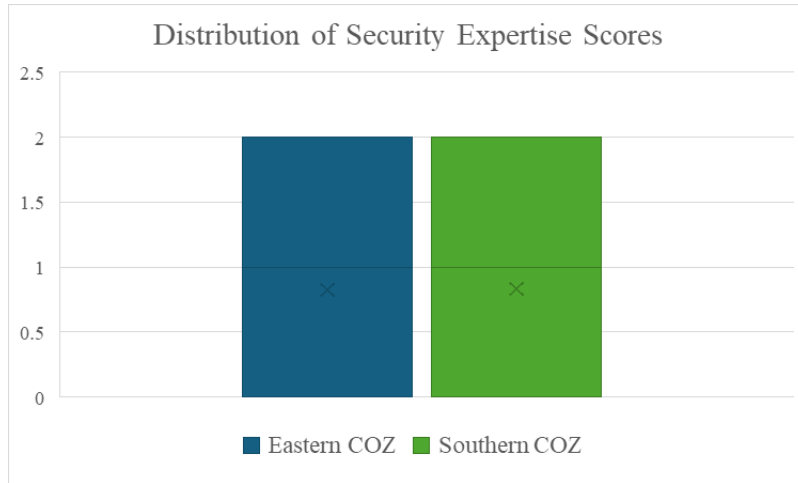


Figure 3. Distribution of security expertise scores.

C. Risk Assessment Behavior

In order to gauge what participants pay attention to when determining the legitimacy of online communication, as noted in the experimental design, we presented participants with a list of common security indicators: https, the lock icon, certificates, privacy policies, and professional appearance. Participants were asked to indicate which of the security indicator(s) plays a role in their decision-making process.

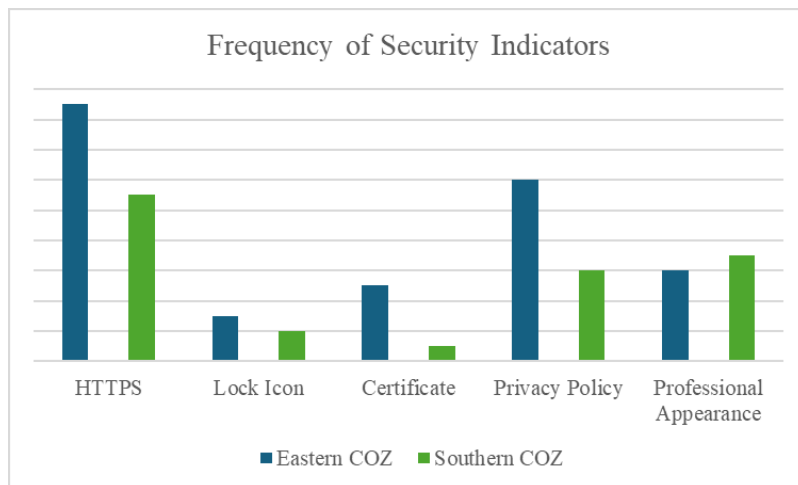


Figure 4. Frequency of security indicators.

Of particular interest are the insights measurable from the proportion of participants who indicated that professional appearance was instrumental in determining legitimacy online. Although professional appearance does not rank among the top indicators of security selected by the participants polled in this experiment, the importance of this factor cannot be underestimated given its position outweighing certificates and the presence of the lock icon, two hallmark features of secure information transfer. The significance of this matter in terms of legitimacy and Latinization will be revisited in the discussion section.

D. Phishing Resilience

For the purposes of this experiment, phishing resilience is defined as the ability to accurately identify legitimate, trustworthy emails, as well as phished, untrustworthy emails. The principal task of this experiment consisted of presenting participants with ten legitimate emails and three illegitimate (phishing) emails in an online survey. Participants were asked to read each email and indicate whether or not the message seemed trustworthy or untrustworthy. In order to quantify the results, a resilience score was calculated by computing the fraction of emails classified as legitimate over the total number of legitimate emails presented and the fraction of emails classified as illegitimate (phished) over the total number of illegitimate emails presented. To calculate a resilience score concerning participants' perceptions of the red herring emails, the fraction of classifications indicating legitimacy over the total number of red herring emails was computed. In order to identify differences in phishing resilience across the Southern and Eastern COZ, a series of one-way ANOVA test were conducted to measure resilience in identifying legitimate emails, illegitimate emails, and the perception of Latinization represented by the red herring emails.

Results of the one-way ANOVA test show that levels of phishing resilience with regards to identifying legitimate emails were not statistically significant between the Southern and Eastern COZ (p-value 0.15).

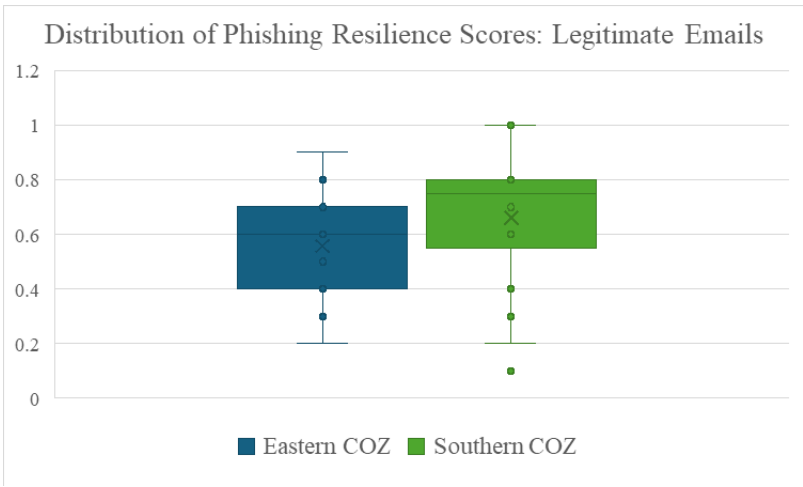


Figure 5. Distribution of phishing resilience scores: legitimate emails.

Results of the one-way ANOVA test show that levels of phishing resilience with regards to identifying illegitimate emails were not statistically significant between the Southern and Eastern COZ (p-value 0.34).

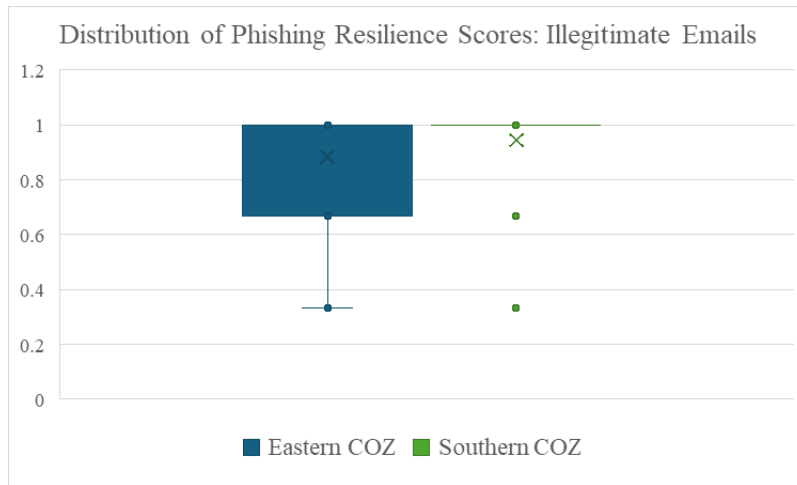


Figure 6. Distribution of phishing resilience scores: illegitimate emails.

Results of the one-way ANOVA test show that levels of phishing resilience with regards to identifying legitimate emails amongst the red herring group were statistically significant between the Southern and Eastern COZ (p-value 0.01). Here, the red herring emails were identified (correctly) as legitimate at significantly higher rates in the Eastern COZ than in the Southern COZ.

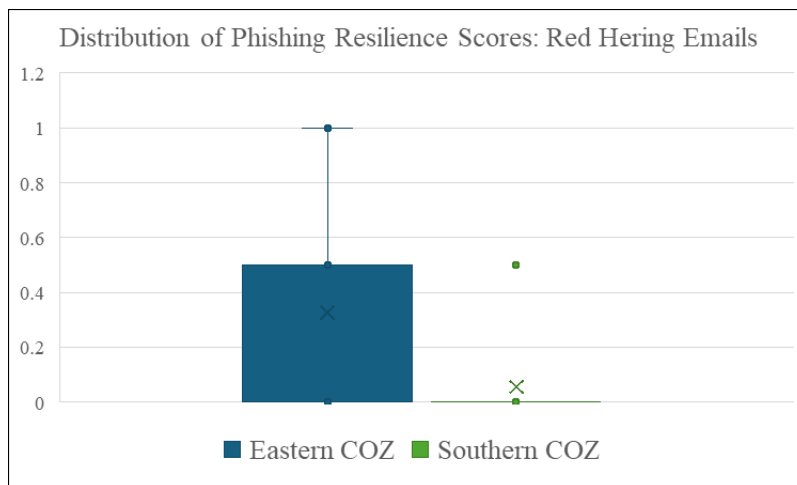


Figure 7. Distribution of phishing resilience scores: red herring emails.

Central to the concerns of this work is the perception of Latinization and its effect on determining legitimacy in digital written communication. To this end, a list of common security indicators was presented to participants in order to measure the importance of professional appearance in determining legitimacy with regard to writing style and usage. An additional one-way ANOVA test was conducted on the results obtained from those participants indicating professional appearance as a key indicator of security and legitimacy in digital written communication. The results of this test show that there is a significant difference in the rates of identifying Latinized red herring emails as safe, trustworthy, and legitimate messages (p-value 0.03) between the Southern and Eastern COZ groups. Among these participants pertaining to the Southern COZ group, Latinized

red herring emails were identified as illegitimate and untrustworthy in 100% of cases. In contrast, those participants pertaining to the Eastern COZ group deemed Latinized red herring emails as illegitimate and untrustworthy in 63% of cases.

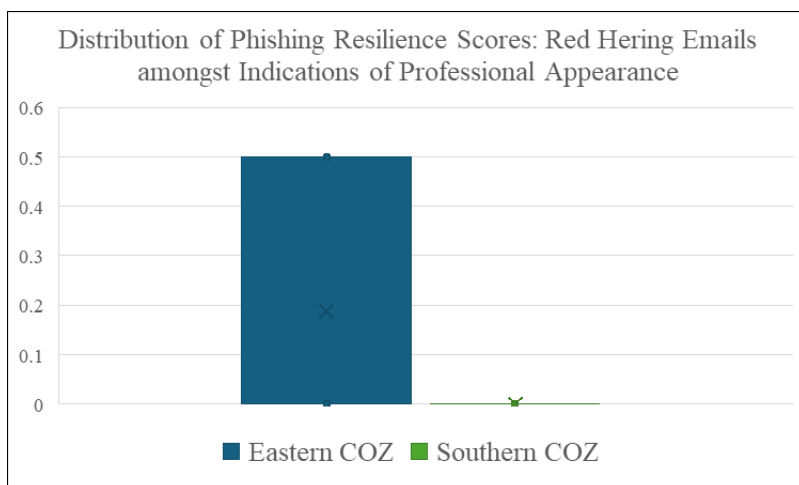


Figure 8. Distribution of phishing resilience scores: red herring emails amongst indications of professional appearance.

V. Discussion and Implications

The primary goal of this work is to address the perception of non-nativeness in online written communication and seek to understand its impact on human users, who ultimately constitute the final line of defense against common e-crime such as phishing. The primary goal of the experiment is to determine how factors such as demographics, knowledge, skills, and risk assessment relate to phishing resilience. This experiment was conducted within the eastern and southern portions of the Cyrillic Orthographic Zone in order to address the topic of non-nativeness and its relation to phishing resilience within linguistic communities beyond English-speaking communities, on which related works are typically focused.

The results of this study provide responses to the research questions posed in Section I:

RQ1: For the purpose of this work, the meaning of style and usage errors was expanded beyond the traditional anglophone understanding to include the presence of Latinization. The results of this work show that Latinization can be more closely correlated with online risk perception in the Southern Cyrillic Orthographic Zone versus in the Eastern Cyrillic Orthographic Zone. More specifically, Bulgarian participants were significantly more likely than Russian participants to react negatively to instances of Latinized text and interpret these instances as indicators of insecurity online.

RQ2: The results obtained in this experiment can be further correlated to the findings of a 2026 study determining the rates of Latinization amongst eastern Slavic languages on social media writing [Sme26]. In comparison to the predominately Russian-speaking region of the Eastern Cyrillic Orthographic Zone, instances of Latinized digital text are significantly more common in the southern, Balkan regions. The results of this experiment warrant a discussion about exposure to Latinization and the perception thereof in different regions of the Cyrillic Orthographic Zone. In countries within the Southern Cyrillic Orthographic Zone like Bulgaria, the heightened presence of Latinization (and its inherent non-nativeness) has resulted in a more widespread, public discussion on literary changes [Kir18]. Exposure to Latinization in the Eastern Cyrillic Orthographic Zone is far less common and, as a result, internet users may interpret the presence of Latinization in different ways.

RQ3: As discussed previously, one must consider the significance of exposure to Latinization in order to interpret how internet users in different countries react to its presence. Seeing that participants in Bulgaria were significantly more likely to associate instances of Latinization with insecurity and risk online, we can correlate heightened exposure to this literary phenomenon with an increased degree of perceived illegitimacy. Conversely, participants from Russia, who are significantly less exposed to instances of Latinization, were significantly less likely to associate Latinization with inherent illegitimacy. Given the lack of exposure to Latinization in the Eastern Cyrillic Orthographic Zone, one must also consider alternative interpretations provided that there is no well-documented public discussion around Latinization and literary change in the Eastern Cyrillic Orthographic Zone as in the Southern Cyrillic Orthographic Zone. A lack of exposure to Latinization may elicit altogether different reactions from internet users in the Eastern Cyrillic Orthographic Zone such as a failure to register the written text as Russian or Ukrainian language. One of this study's limitations is the inability to interview participants while they make decisions. In such possibilities, internet users may superficially mistake the text as another related language officially written in Latin script such as Czech or Slovak and immediately focus on other more-commonly known indicators of security such as the presence of HTTPS, the lock icon, or the overall professional appearance of the website. It is important to also highlight the role of political discourse in participants' decision making. Latinization represents a current social issue in the Balkans while there are significantly less debate and literature addressing this topic in Russian-speaking territories further east.

This work also serves as a basis for future research to explore the policy implications of these findings, particularly within the European context. European countries continue to strengthen cybersecurity through initiatives such as the European Union Agency for Cybersecurity guidelines, the European Commission's Cybersecurity Strategy, and the implementation of certain directives such as NIS2. As illustrated in this work, linguistic and cultural factors can play a role in determining how internet users perceive certain risks online and therefore present an opportunity to expand upon existing cybersecurity awareness programs. The findings presented in this study suggest that differences between writing traditions in Europe and varying levels of exposure to non-native digital communication can shape perceptions of legitimacy and risk online. Future work could therefore examine how cybersecurity policies, public awareness campaigns, and educational materials can be adapted to account for Europe's particularly diverse linguistic environment, which involves digital text production in a variety of different scripts.

VI. Conclusion

This study extends the understanding of phishing resilience by situating perceptions of non-nativeness within a more globally inclusive understanding of cyber risk perception. Moving beyond the predominantly Anglocentric focus of prior research on risk perception, the findings demonstrate that linguistic cues associated with determining legitimacy online do not function as universal indicators of risk. This work addresses the role of communication amongst internet users whose native languages do not use Latin script, particularly Cyrillic in Eastern Europe.

While traditional phishing markers such as urgency, impersonality, and grammatical inconsistency remain relevant, this study shows that what constitutes a "suspicious" writing style is not globally consistent. Regionally specific factors such as Latinization and digraphia are interpreted in users' linguistic environments and influenced by levels of exposure to non-traditional writing styles. The results further highlight that exposure plays a decisive role in shaping cyber risk perception. Participants from the Southern Cyrillic Orthographic Zone, who encounter Latinization more frequently, were more likely to associate it with illegitimacy, whereas participants from the Eastern Zone demonstrated alternative interpretations of the same feature. Even within the Cyrillic Orthographic Zone there is not a single approach to evaluating linguistic cues.

These findings carry important implications for both research and practice. From a research perspective, they emphasize the need for more cross-cultural and cross-linguistic studies that capture not only the diversity of languages utilized by global internet users, but the range of circumstances which may or may not be applicable to internet users belonging to certain linguistic communities. From a practical standpoint, they suggest that anti-phishing strategies, user education programs, and automated detection systems can be expanded to include key behavioral dynamics of certain communities for whom the expansion of digital text production has incurred significant changes to traditional writing practices.

References

- [BPC11] Mark Blythe, Helen Petrie, and John A. Clark. F for fake: Four studies on how we fall for phish. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.
- [BVK⁺15] Christian Bentz, Annemarie Verkerk, Douwe Kiela, Felix Hill, and Paula Buttery. Adaptive communication: Languages with more non-native speakers tend to have fewer word forms. *PLOS ONE*, 10(6):e0128254, 2015.
- [DDMP21] Giuseppe Desolda, Francesco Maria Donini, Valentina Marrella, and Cesare Pappalardo. Human factors in phishing attacks: a systematic literature review. *ACM Computing Surveys (CSUR)*, 54(8):1–35, 2021.
- [Fur07] Steven Furnell. Phishing: Can we spot the signs? *Computer Fraud & Security*, 2007(3):10–15, 2007.
- [GC21] Jayati Dev Marthie Grobler DongInn Kim Sanchari Das Gopavaram, Shakthidhar Reddy and L. Jean Camp. Cross-national study on phishing resilience. In *Proceedings of the Workshop on Usable Security and Privacy (USEC)*, 2021.
- [Geo17] S. Georgieva. From communicative competence of the teacher to communicative competence of the student. pages 407–418, 2017.
- [GF24] Shih-Wei Guo and Yao-Chung Fan. X-Phishing-Writer: A framework for cross-lingual phishing e-mail generation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(7):1–34, 2024.
- [KB16] Timothy Kelley and Bennett I. Bertenthal. Attention and past behavior, not security knowledge, modulate users’ decisions to login to insecure websites. *Information & Computer Security*, 24(2):164–176, 2016.
- [Kir18] Lyudmila Kirova. Săvmestna upotreba na kirilitsa i latinitsa v bălgarski tekst – zashto i koga. *Slavica Lodziensia*, 2:219–231, 2018.
- [MH20] John McAlaney and Peter J. Hills. Understanding phishing email processing and perceived trustworthiness through eye tracking. *Frontiers in Psychology*, 11:1756, 2020.
- [MXR23] Elaheh Mehrabi, Tianhao Xu, and Prashanth Rajivan. SpearSim-V2: Synthetic task environment for evaluating attacker behaviors. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 67, Los Angeles, CA, 2023. SAGE Publications.
- [OMA20] Enrique Orduña-Malea and Isidro F. Aguillo. Brief history of top-level domains and challenges for information professionals. *Anuario ThinkEPI*, 14, 2020.
- [PBP⁺16] Kathryn Parsons, Marcus Butavicius, Malcolm Pattinson, Calic Dragana, Agata McCormac, and Cate Jerram. Do users focus on the correct cues to differentiate between phishing and genuine emails? *arXiv preprint arXiv:1605.04717*, 2016.
- [Ria07] Vladimir Riabov. Smt p (simple mail transfer protocol). *The Handbook of Computer Networks*, pages 388–406, 2007.
- [Sch12] Deborah Schaffer. The language of scam spams: linguistic features of “Nigerian fraud” e-mails. *ETC: A Review of General Semantics*, pages 157–179, 2012.
- [Sme26] William Smeal. Establishing rates of latinization in the slavic cyrillic orthographic zone: A corpus-based analysis. *Diss. Indiana University*, 2026.
- [Sun10] Hong-mei Sun. A study of the features of Internet English from the linguistic perspective. *Studies in Literature and Language*, 1(7):98, 2010.

- [UDKC20] Ploy Unchit, Sanchari Das, Andrew Kim, and L Jean Camp. Quantifying susceptibility to spear phishing in a high school environment using signal detection theory. In *International Symposium on Human Aspects of Information Security and Assurance*, Cham, 2020. Springer International Publishing.
- [V*24] Sophomore T. Vacalares et al. Bank emails: The language of legit and scam. *International Journal of Research and Review*, 11:192–203, 2024.
- [Wol04] Robert M. Wolk. The effects of english language dominance of the internet and the digital divide. In *2004 International Symposium on Technology and Society*. IEEE, 2004.
- [WR12] Mark Gupta Wang, Jingguo and H. Raghav Rao. Research article phishing susceptibility: An investigation into the processing of a targeted spear phishing email. *IEEE Transactions on Professional Communication*, 55(4):345–362, 2012.

Appendix 1. POST EXPERIMENTAL SURVEY QUESTIONS (BULGARIAN)

- 1) Какво е фишинг? (изберете всички подходящи варианти) What is phishing? (Select all applicable answers.)
- 2) Какво е целта на сертификат X.509? (Изберете всички подходящи отговори) What is the purpose of an X.509 certificate for websites?
- 3) Инжекция SQL се използва за: An SQL injection is a technique used to:
- 4) Каква е разликата между пасивна и реактивна система за откриване на проникване (Intrusion Detection System)? What is the difference between a passive and reactive Intrusion Detection System?
- 5) Без никакви промени в настройките на уеб сервера, какво може да бъде мотивацията да се затвори порта 80? Without any other changes in the default settings of a web server, what can be a motivation to close port 80?
- 6) Колко компютърни езици знаете? (без да включвате HTML)? How many computer programming languages do you know (excluding HTML)?
- 7) Колко години опита имате в сферата на мрежова сигурност? How much experience in network operation and security do you have?
- 8) Какви инструменти за мрежова сигурност Вие регулярно използвате? (Може да изберете няколко варианти) What information and network security tools do you use regularly? (Select all applicable answers.)
- 9) Кои от следващите показатели Вие използвате за да решите дали е безопасно да въведеш паролата Ви в определен сайт? (Изберете всички подходящи отговори) Which of the following indicators do you use to decide if it is safe to enter your username and password on a particular site?

Appendix 2. POST EXPERIMENTAL SURVEY QUESTIONS (RUSSIAN)

- 1) Что такое фишинг? (Выбирайте все подходящие ответы.) Привет мир What is phishing? (Select all applicable answers.)
- 2) Какова цель сертификата X.509 для интернет страницы? (Выбирайте все подходящие ответы.) What is the purpose of an X.509 certificate for websites?
- 3) SQL инъекция – это метод для: An SQL injection is a technique used to:
- 4) В чем разница между пассивной и активной системой обнаружения несанкционированного доступа? What is the difference between a passive and reactive Intrusion Detection System?
- 5) Без каких-либо других изменений в настройках веб-сервера, какая может быть мотивация для закрытия порта 80? Without any other changes in the default settings of a web server, what can be a motivation to close port 80?
- 6) Сколько компьютерных языков Вы знаете? (не включая HTML)? How many computer programming languages do you know (excluding HTML)?
- 7) Есть ли у Вас опыт в области сетевых операций и сетевой безопасности? How much experience in network operation and security do you have?
- 8) Какие информации и инструменты сетевой безопасности вы регулярно используете? (выбрать все подходящие ответы) What information and network security tools do you use regularly? (Select all applicable answers.)
- 9) Какие из следующих указателей Вы ищете, чтобы понять, безопасно ли будет переходить в определённый сайт? Which of the following indicators do you use to decide if it is safe to enter your username and password on a particular site?