# CAPITALISATION ERRORS IN WRITTEN WORKS OF LEARNERS OF LATVIAN AS A FOREIGN LANGUAGE

**Antra Kļavinska**
Rezekne Academy of Technologies, Latvia

**Abstract.** *Capitalisation in writing is usually determined by tradition. Different written languages can have their own grammatical, conceptual or stylistic capitalisation rules. Orthographies exist which do not have the division into capital and small letters.*
*The aim of the article is to find out what problems with capitalisation foreign students in Latvian higher education institutions have during the acquisition of writing skills in Latvian as a foreign language. The research source are the essays written by learners of the Latvian language (foreign students studying in Latvian higher education institutions): the data of the Latvian language learner text corpus being created in the Institute of Mathematics and Computer Science of the University of Latvia were used. The requirements for the acquisition of capitalisation in the context of language learner competences are analysed in the study; the most typical capitalisation errors and possible reasons for them are analysed; and the author's practical experience teaching the Latvian language to foreign students is revealed.*
**Keywords:** *Latvian as a foreign language, learner corpus, capitalisation, error analysis.*

## Introduction

Latvian as a foreign language can be studied at higher education institutions and language teaching centres both in Latvia and in more than 20 higher education institutions abroad (in Europe, the United States, China, etc.). Since Latvia joined the EU in 2004, the interest in learning Latvian in Latvia as a foreign language has increased; interest primarily comes from foreigners who have arrived to study or work. Latvian higher education institutions have full-time students as well as exchange students from almost every place of the world (Laizāne, 2019).

It is only natural that students make different mistakes while learning a foreign language. In applied linguistics, a distinction is made between the terms "mistake" and "error". For example, Ellis (1997, 17) states that mistakes are caused by carelessness and cannot be considered an erroneous interpretation of the grammatical rules of the second language. Mistakes can be self-corrected when attention is called. Whereas errors occur due to lack of knowledge, and thus the learner cannot self-correct them. In this article, the term "errors" is used to describe the analysis of the errors made by the language learners in tests (essays).

"The aims of the studies regarding error analysis can be summarized as follows: Error analysis identifies the strategies that language learners use. It looks for the answer of the question 'why do learners make errors?' It determines the common difficulties in learning and helps teachers to develop materials for remedial teaching" (Erdoğan, 2005, 269).

More and more new digital resources are created nowadays, which are useful for foreign language teachers and specialists in applied linguistics. One such resource is error-tagged learner corpora. Considering the popularity of the English language as an intermediary language, the most attention has been given to the creation of English language learner corpora; however, in the recent decades, learner corpora have been created for other languages as well (Granger, 2008). Sylviane Granger names core components for learner corpus research: corpus linguistics, linguistic theory, second language acquisition, and foreign language teaching (Granger, 2009).

To enable the study of the specific nature of learning Latvian as a second language and as a foreign language, as well as the analysis of errors made by Latvian language learners and the development of corpus-based teaching materials, Latvian language learner corpora have also been created: a corpus of texts collected from successfully passed State Language Proficiency Testing, which is used to evaluate a person's state language proficiency level (VVP, 2018); and the Learner Corpus of the Second Baltic Language (ESAM, 2014). The Institute of Mathematics and Computer Science at the University of Latvia is creating a corpus of school student essays, which includes more than 400 essays from 12th grade Latvian language exams. These are the works of students from secondary schools, national minority schools and state grammar schools with the Latvian language of instruction (Pārspriedumi, 2018-2021) and the Latvian language learner corpus (LaVA, 2018-2021). The LaVA corpus includes the works of foreign students in the Latvian higher education institutions who have been learning Latvian as a foreign language for the first or second semester. The texts are subjected to automated morphological analysis with learner errors marked manually. The completion of the project is planned for August 2021; the planned corpus size is 1000 student works (essays); the length of each text is approximately 100 words (approximately 100 thousand words in total). The planned amount of data has already been collected, and automated error tagging and manual reviewing is in progress.

When teaching Latvian to foreign students, the author of the article has noticed several typical errors students make when learning the Latvian language. One of the spelling errors is incorrect capitalisation. The aim of this article is to find out what problems with capitalisation foreign students in Latvian higher education institutions have when acquiring the writing skills in Latvian as a foreign language. The research source are the essays written by learners of the

Latvian language (foreign students studying in Latvian higher education institutions) – the data of the Latvian language learner text corpus (LaVA), as well as some works of foreign students (level A1, 21 essay, length of each essay ~100 words, academic year 2018/2019). The data collection methods used were corpus linguistics methods as well as manual collection of text excerpts; quantitative and qualitative analysis of capitalisation errors was performed in the study.

The problem questions raised in the study:

1. What is the ratio of capitalisation errors and other types of errors made by the language learners?
2. What are the requirements for the acquisition of capitalisation in the context of learners of Latvian as a foreign language?
3. What are the most typical capitalisation errors in the written works of learners of Latvian as a foreign language?
4. What are the causes of capitalisation errors?

**Capitalisation Errors vs. Other Types of Errors**

Capitalisation in writing is usually determined by tradition. Different written languages can have their own grammatical, conceptual or stylistic capitalisation rules. Orthographies exist that do not have the division into capital and small letters (for example, the Chinese script, Devanagari script); whereas, for example, in German orthography, capitalisation is used as a marker for all nouns.

In the Latvian language, capitalisation has grammatical meaning when the word is used at the beginning of a sentence and at the beginning of a direct quotation. Whereas capitalisation of proper nouns and compound proper names has a conceptual semantic sense. As for compound names used in the sense of a proper noun, which are expressed using several words, either every word in the name is capitalised or only the first one (Porīte, 1970; Strautiņa & Šulce, 2009). Moreover, some nouns that are considered proper nouns and capitalised in English and other languages are not capitalised in Latvian, i.e., ethnonyms, such as *latvietis* 'Latvian' and linguonyms, such as *latviešu valoda* 'Latvian language', *runāt latviski* 'to speak Latvian'.

During the creation of text corpora and error tagging, a taxonomy of the errors found in the written works of Latvian language learners with the classification of error types and subtypes was created: 1) spelling errors (upper / lower case letter, diacritics, words spelled separately / together, missing letters, redundant letters, other spelling errors); 2) punctuation errors (missing punctuation, redundant punctuation, incorrect punctuation); 3) grammatical errors (incorrect word form (such as inflection, gender, number, definite/indefinite ending, tense, person), derivation, morphophonemic consonant alternation); 4) syntactic errors (word order, redundant word, missing word); 5) lexical errors

(meaning, compliance, readability, collocation); 6) unclear text (Dargis, Auzina, & Levane-Petrova, 2018, 4111-4112).

Since the LaVA corpus and the error tagging process is still incomplete, it is impossible to perform a comprehensive quantitative error analysis in order to determine the frequency of capitalisation errors versus other types of errors. However, for example, in the study on the results of the state language proficiency test (Auziņa et al., 2019), a quantitative and qualitative analysis of language errors was performed using the state language proficiency (VVP) corpus data (146 806 text units from written state language proficiency tests). It was concluded that the most prevalent errors are spelling errors (37 %) followed by punctuation (18 %), inflection and word formation errors (18 %), combined errors (spelling and inflection, lexical un spelling, etc.) (12 %), syntactic errors (8 %), lexical errors (4 %), and unclear text (4 %) (Auziņa et al., 2019, p. 82). The most prevalent spelling errors are missing or redundant diacritics, distinction between vowels and diphthongs, rendering of proper names. Analysis of capitalisation in the study shows that such errors are not many. This could possibly be explained by the fact that the native language of the test takers was usually Russian or another language where the spelling rules regarding capitalisation and proper nouns are quite similar to the principles of the Latvian language. Most frequently, the language learners made errors in the letter (one of the tasks in the test) by not capitalising the pronoun forms *Tu, Jūs,* and *Tavs*. Sometimes the first letter of the sentence was written with a small letter, although this is probably a careless mistake rather than a true error (Auziņa et al., 2019, 90).

Capitalisation errors are one of the most frequent types of errors in the Latvian Language and Literature centralised examination works (the exam is taken by students who speak Latvian as a native language as well as a second language (Špūle *et al.,* 2007, 38)), and one of the most frequent linguistic faults in everyday practice is that the main problems with capitalisation are observed in the names of different organisations. Although normative materials describe the general capitalisation principles, many problem cases still exist because language users often have no understanding of symbolic and direct names, as well as whether the word is used as a proper noun (Project "Izplatītākās valodas nepilnības un ieteikumi to novēršanai" ('Common language gaps and recommendations for their prevention') by the University of Latvia). Thus, capitalisation errors are characteristic not only of learners of Latvian as a foreign or second language but also learners and speakers of Latvian as a native language.

## Capitalisation in the Context of Competencies of a Language Learner

The description of Latvian language learner competencies shows that the requirements for capitalisation differ depending on the level of language proficiency. At level A1, the language learners are expected to spell familiar words and write them down using capital and small letters in cursive and in print. The description of the intermediate level of language proficiency (B1, B2) states that the language learners have to follow specific rules in business writing, including capitalisation (Šalme & Auziņa, 2016a).

Whereas the improvement of orthography skills at the highest level (C1 and C2) presupposes correct spelling of proper nouns, focusing on capitalisation of proper nouns. The description also specifies that at the highest level of language proficiency, focus is given to the rules of capitalisation (Šalme & Auziņa, 2016b).

Thus, the description of the Latvian language proficiency levels provides that at the basic level (A1 and A2) and at the intermediate level (B1, B2) no special attention is given to capitalisation in the learning process; this topic is acquired in detail only at the highest level (C1, C2). However, considering that the basic and advanced level still include topics which require the use of capitalisation (for example, countries and cities, business writing, etc.), attention should be given to this aspect of spelling, especially in the cases specific for the Latvian language.

## Most Typical Capitalisation Errors

Capitalisation errors are generally described as underuse and overuse of capital letters (Söderlind, 2008). In this study, capitalisation errors have been grouped into grammatical and semantic context, i.e., incorrect spelling of appellatives and proper nouns. The following errors have been found concerning the expression of grammatical meaning in the written works of Latvian language learners: missing capitalisation at the beginning of a sentence and unnecessary capitalisation (in the middle of a sentence) – see Table 1. In the table, the first number in the parentheses shows the number of instances in the Latvian language learner corpus, but the second number shows the number of essays containing such an error; these examples also include other orthographic errors made by students; however, these are not analysed.

*Table 1* **Most Fequent Capitalisation Errors in Appellatives** *(LaVa corpus data, 2019)*

| Missing capitalisation at the beginning of a sentence | Unnecessary capitalisation (in the middle of a sentence) |
|---|---|
| • *es* 'I' (13/12)<br>• *man* 'to/for me' (14/10)<br>• *un* 'and' (6/6)<br>• *mana* 'my' (4/3)<br>• *mani* 'me' (3/3)<br>• *viņa* 'she' (4/3) | • *Patik* 'like' (17/7)<br>• *Universitātē* 'at university' (16/15)<br>• *Patīk* 'like' (15/5)<br>• *Sauc* 'call [they call me = my name is'] (11/11)<br>• *Es* 'I' (10/9)<br>• *Mana* 'my' (4/4)<br>• *Piens* 'milk' (4/4)<br>• *Alus* 'beer' (4/4)<br>• *Pusdienās* 'for lunch' (4/4) |

Capitalisation errors in the semantic sense (in proper nouns) can also be divided into two groups: 1) missing capitalisation in toponyms, for example, *(no) vacijas* (correct: *no Vācijas* 'from Germany'), ergonyms, for example, *stockpot* (correct: *Stockpot*), in compound names, for example, *Rīgas stradiņa universitāte* (correct: *Rīgas Stradiņa universitāte* 'Rīga Stradiņš University'), street names, for example, *stabu ielā* (correct: *Stabu ielā* 'in Stabu street'); 2) unnecessary capitalisation in ethnonyms, for example, *Amerikanis* (correct: *amerikānis* 'American') and linguonyms, for example, *Angliski* (correct: *angliski* 'in English') – see Table 2.

*Table 2* **Most Frequent Capitalisation Errors in Proper Nouns** *(LaVa corpus data, 2019)*

| Missing capitalisation in proper nouns | Unnecessary capitalisation in proper nouns |
|---|---|
| - *Rīgas stradiņa universitāte* 'Rīga Stradiņš University' (6/6)<br>- *(no) vacijas* '(from) Germany' (5/5)<br>- *stockpot* (4/4)<br>- *(no) latvijas* '(from) Latvia' (3/3)<br>- *vacijā* 'in Germany' (3/3)<br>- *stabu (ielā)* '(in) Stabu (street)' (2/2) | - *Angliski* 'in English' (6/5)<br>- *Latviski* 'in Latvian (5/4)<br>- *Amerikanis* 'American' (1/1)<br>- *Eiropieši* 'Europeans' (1/1)<br>- *Francūziete* 'French' (1/1) |

The following is the analysis of capitalisation in the works of specific students (21 essay, topic – "My studies"), which were written in academic year 2018/2019 by A1 level students at the end of the semester. Student profile: These are regular students acquiring a full education programme in Latvia and Erasmus+ exchange students. The authors of the essays have stated the following native

languages: Turkish (7), French (4), Ukrainian and Russian (3), Ukrainian (2), Hindi (2), Punjabi (2), Chichewa (1); knowledge of foreign languages: English (21), Polish (3), German (2), Russian (2), Spanish (2), Hindi (2), Turkish (1).

First, a quantitative analysis of the errors was performed – see Table 3.

*Table 3* **Number of Capitalisation Errors in the Same Essay**

| Number of errors in the same essay | Number of essays |
| --- | --- |
| No errors | 6 |
| One error | 5 |
| Two errors | 3 |
| Three errors | 1 |
| Four errors | 2 |
| Five errors | 2 |
| 11 errors | 1 |
| 21 error | 1 |

The works of the students demonstrate very different levels of performance. Six essays contain no capitalisation errors (the authors are students with different native languages – Turkish, French, Russian, Ukrainian); whereas one essay shows chaotic capitalisation (21 error; the author's native language is Punjabi; knowledge of foreign languages – Hindi, English)): *Es runaju angliski, Punjabiski, Hindiski un Latviski Mazliet. Man Patik macīties Rēzekne Jo Visi Skolotāji ir ļoti noderigi un jauki. Man Patik rēzeknes Pilsēta jo man ļoti Patik Kultura daba un Cilvēki.* (Correct: *Es runāju angliski, pendžabiski / pendžabiešu valodā, hindiski un mazliet latviski. Man patīk mācīties Rēzeknē, jo visi skolotāji ir ļoti noderīgi un jauki. Man patīk Rēzeknes pilsēta, jo man ļoti patīk kultūra, daba un cilvēki.* 'I speak English, Punjabi language, Hindi and a little Latvian. I like studying in Rezekne because all the teachers are very useful and nice. I like the city of Rezekne because I like the culture, nature and the people very much.')

Qualitative error analysis shows the same errors as found in the LaVa corpus. Missing capitalisation is found most frequently in toponyms (names of countries and cities) – 5 errors in total, e.g., *Vins ir no latvijas* (correct: *Viņš ir no Latvijas* 'He is from Latvia'). There are fewer errors at the beginning of a sentence (2 errors), e.g., *es esmu no Indijas* (correct: *Es esmu no Indijas.* 'I am from India.'), and in compound ergonyms (2 errors), e.g., *Es esmu students en Rezekne technologiju akademija* (correct: *Es esmu students Rēzeknes Tehnoloģiju akadēmijā.* 'I am a student at Rezekne Academy of Technologies').

A more common problem is unnecessary capitalisation: in linguonyms (14 errors), e.g., *Es runaju Turčiski, Angliski un Latviski mazliet* (correct: *Es runāju turciski, angliski un mazliet latviski.* 'I speak Turkish, English and a little Latvian'); in the middle of a sentence (in nouns, verbs, adjectives (13 errors)),

e.g., *Latviešu valoda ir Sarezgita* (correct: *Latviešu valoda ir sarežģīta.* 'Latvian language is difficult.'); in ethnonyms (3 errors), e.g., *Vinš ir Korejietis* (correct: *Viņš ir korejietis* 'He is Korean').

## Possible Causes of Errors

The causes of errors are categorised into interlingual transfer and intralingual transfer. Interlingual transfer can be either positive or negative, which is characterised by the use of native language structures during the acquisition of a foreign language. Interlingual errors are the result of negative interlingual transfer (Ellis, 1997, 19). Intralingual errors are classified into four categories including overgeneralization, ignorance of rule restrictions, incomplete application of the rules, and false concepts hypothesised or semantic errors (Richards, 1974, 120).

Capitalisation traditions are similar in many European languages, which is why in the acquisition of Latvian spelling it can generally be seen as positive interlingual transfer. However, there are several exceptions – error analysis has also discovered cases of negative transfer, i.e., ethnonyms and linguonyms, which, contrary to the Latvian language spelling, are considered proper nouns and capitalised in many languages.

As the author's observations show, regardless of the previous knowledge of the English language, which is used as an intermediary language in Latvian language classes, students from India and Pakistan have the most difficulty with capitalisation because their first languages use Devanagari script, which has no division into capital and small first letters.

Capitalisation errors in compound proper names can be interpreted as a result of intralingual transfer because the different rules pose problems even for native language learners. Moreover, considering that in online communication even native speakers do not use capital letters or use them in a chaotic manner (Urbanoviča, n. d.), ignoring capitalisation rules could gradually become a universal spelling simplification phenomenon.

It is possible that some of the capitalisation errors found were really just mistakes, for example, missing capitalisation at the beginning of a sentence or capitalisation of appellatives mid-sentence. Sometimes it was difficult to tell from the handwriting whether the student meant to write a capital or a small letter.

To ensure more successful acquisition of capitalisation rules when teaching Latvian to foreign students, attention needs to be given gradually, starting from the beginner's level, to the experience and habits of the language learners (to prevent errors caused by interlingual transfer) as well as the specific rules of the Latvian language (to prevent errors caused by intralingual transfer).

## Acknowledgements

## References

Auziņa, I., Kļava, G., Lazareva, A., Levāne-Petrova, K., Mūrniece-Buļeva, B., Pāvulēna, S., & Semjonova, A. (2019). *Latviešu valodas prasmes kvalitāte: valsts valodas prasmes parbaudes rezultāti.* Rīga: Latviešu valodas aģentūra 2019. Retrieved from: https://valoda.lv/wp-content/uploads/2020/04/Latvie%C5%A1u-valodas-prasmes-kvalit%C4%81te-preview.pdf

Dargis, R., Auzina, I., & Levane-Petrova, K. (2018). The Use of Text Alignment in Semi-Automatic Error Analysis: Use Case in the Development of the Corpus of the Latvian Language Learners. *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*, 4111 – 4115. Retrieved from http://www.lrec-conf.org/proceedings/lrec2018/pdf/933.pdf

Ellis, R. (1997). *Second Language Acquisition.* Oxford: Oxford University Press.

Erdoğan, V. (2005). Contribution of Error Analysis to Foreign Language Teaching. *Mersin University Journal of the Faculty of Education, Volume 1*, Issue 2, December 2005, 261-270 Retrieved from: https://www.researchgate.net/publication/265148899_Contribution_of_Error_Analysis_to_Foreign_Language_Teaching

Granger, S. (2008). Learner corpora. In Lüdeling, A., Kytö, M. (eds.), *Corpus linguistics: An International Handbook (*259-275*).* Berlin, New York: Walter de Gruyrter.

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In A. Karin (Ed.), *Corpora and Language Teaching* (13-32). Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl.33.04gra

Laizāne, I. (2019). Latviešu valoda kā svešvaloda: lingvodidaktikas virziena attīstība Latvijā un ārpus tās. *Promocijas darbs.* Liepāja: Liepājas Universitāte. Retrieved from https://www.liepu.lv/uploads/dokumenti/prom/Disertacija_Laizane_2019.pdf

LaVA. (2018-2021). *Latviešu valodas apguvēju korpuss.* Latvijas Universitātes Matemātikas un informātikas institūts. Retrieved from: http://lava.korpuss.lv/

Latvijas Universitāte (LU). (n.d.). Projekts *Izplatītākās valodas nepilnības un ieteikumi to novēršanai.* Retreived from: https://www.lu.lv/filol/valoda/2_sak.htm

ESAM. (2014). *Otrās baltu valodas apguvēju korpuss.* http://esam.korpuss.lv/

Pārspriedumi. (2018-2021). *Skolēnu pārspriedumu korpuss.* LU MII, Liepājas Universitāte, Rēzeknes Tehnoloģiju akadēmija. http://www.korpuss.lv/id/P%C4%81rspriedumi

Porīte, T. (1970). *Lielie burti mūsdienu latviešu valodā.* Rīga: Zinātne.

Richards, J. C. (1974). *Error Analysis: Perspective on Second Language Acquisition.* London: Longman Group Ltd.

Strautiņa, V., & Šulce, Dz. (2009). *Latviešu valodas pareizruna un pareizrakstība.* Riga: RaKa.

Söderlind, E. (2008). *Capital mistakes: The analysis of mistakes in the written production of advanced Swedish ESL learners at university level with focus on the use of capital.* Bachelor thesis. Växjö University. Retrieved from: http://www.diva-portal.org/smash/get/diva2:202574/FULLTEXT01

Šalme, A., & Auziņa, I. (2016a). *Latviešu valodas prasmes līmeņi: pamatlīmenis A1, A2, vidējais līmenis B1, B2*. Rīga: Latviešu valodas aģentūra.

Šalme, A., & Auziņa, I. (2016b). *Latviešu valodas prasmes līmeņi: augstākais līmenis (C1 un C2). Vadlīnijas.* Rīga: Latviešu valodas aģentūra.

Špūle, K., Vanaga, A., Garjāne, B., Dalbiņa, D., & Sniedze, A. (2007). *Skolēnu sasniegumu analīze tekstveidē latviešu valodas un literatūras centralizētajā eksāmenā: situācijas izpēte un ieteikumi.* Rīga. Retrieved from: https://www.visc.gov.lv/lv/petijumi/ssapet1.pdf

Urbanoviča, I. (n.d.). Latviešu valodas īpatnības internetā. *Projekts "Mūsdienu latviešu valoda interneta vidē*. https://www.lu.lv/filol/valoda/ind_3_publ_r_ievads.htm

VVP. (2018). *Valsts valodas prasmes pārbaudes darbu korpuss*. LU MII. http://www.korpuss.lv/id/VVPP