

Original Article

SCRICREA: A Learner Corpus of Creative Writing in Italian as a Foreign Language

¹KATERINA FLOROU, ²IOANNA TYROU

^{1,2}, Department of Italian Language and Literature, National and Kapodistrian University of Athens, Greece.

ABSTRACT: *SCRICREA (from “Scrittura Creativa”) is a learner corpus of creative writing in Italian as a foreign language, developed at the National and Kapodistrian University of Athens. It consists of texts written by Greek-speaking adult learners of Italian in the context of structured creative writing activities. The corpus, comprising over one million words, is organized into thematic sub-corpora, each corresponding to a specific writing task or genre. Its interdisciplinary nature positions SCRICREA both as a tool for linguistic research and as a pedagogical resource for encouraging creativity and language proficiency. This paper presents the motivation behind SCRICREA, its composition, annotation and enrichment strategies, as well as its applications in language pedagogy and future directions.*

KEYWORDS: Creative writing, Learner corpora development, Foreign language analysis, *Scrittura Creativa*.

1. INTRODUCTION

Learner corpora, defined as systematic collections of texts produced by foreign language learners, have become essential resources in applied linguistics and pedagogy [1], [2]. Over the last two decades, corpus-based approaches have been applied not only to error analysis and second language acquisition research but also to pedagogical practices [3], [4]. Incorporating creative writing into learner corpus research represents a novel step, bridging linguistic accuracy and expressive freedom [5], [6], [7]. At the University of Athens, this innovation has taken shape in the SCRICREA corpus, which not only captures interlanguage phenomena but also offers learners meaningful opportunities for artistic expression. As explored in the SCRICREA project, creative writing provides a dynamic framework for second language acquisition, cultivating narrative strategies, self-expression, and intercultural engagement beyond grammatical accuracy [8]. Its pedagogical extension further demonstrates how learner texts can reveal lexical awareness, cross-linguistic reflection, and language difficulties, enabling educators to harness NLP-based tools to improve linguistic precision while sustaining creative fluency [8]. Learner Corpus Research (LCR) has grown substantially over the last twenty years [9], moving from error-tagged corpora to sophisticated, multi-layered datasets that allow for cross-linguistic comparisons and pedagogical applications [10]. Creative writing in foreign language education has also gained traction as a means of combining affective engagement with linguistic development [11]. Integrating the two perspectives, SCRICREA stands as an example of how corpora can simultaneously document learner interlanguage and stimulate expressive freedom.

1.1. CREATIVE WRITING AS A TRANSFORMATIVE PEDAGOGY IN SECOND LANGUAGE AND INTERCULTURAL EDUCATION

Recent scholarship positions creative writing as a multifaceted pedagogical tool in second language education, one that bridges linguistic development with personal and intercultural exploration. [5] demonstrates that creative writing enhances learners’ grammatical accuracy, lexical acquisition, and fluency, while simultaneously promoting cognitive flexibility, self-expression, and critical thinking. This dual impact-linguistic and affective-has led researchers to emphasize its role in cultivating learner motivation and agency.

Central to this approach is the creation of safe, authentic spaces where learners feel empowered to explore identity, voice, and power relations [12], [13]. Such environments support deeper engagement with others’ perspectives and lived experiences [14], transforming writing into a site of personal and social inquiry [15]. When learners write from within, they participate more actively and meaningfully in the learning process [16]. Recent findings further underscore the value of suggestion-based writing prompts, which stimulate creativity, deepen linguistic engagement, and promote authentic self-expression in second language learners [17].

Moreover, creative writing contributes significantly to intercultural education by encouraging students to articulate diverse cultural viewpoints and engage in reflective dialogue. Its integration into language curricula allows educators to address implicit biases, navigate sociopolitical complexity, and foster communicative competence [18]. As Zheng [19] argues, this dynamic practice not only deepens cultural awareness but also promotes critical thinking and personal growth, equipping learners to participate empathetically in global interactions.

2. CORPUS COMPOSITION

2.1. THE LEARNERS

The corpus includes written productions by 339 adult learners, aged 18-52, enrolled in the Italian Language and Literature program at the University of Athens. All participants are native speakers of Greek and most have prior knowledge of other foreign languages (primarily English). Their Italian proficiency ranges from B2 to C1 on the CEFR scale. Approximately one-third already hold prior university degrees. The corpus consists of creative writing texts produced in response to prompts given during coursework. Each weekly unit corresponds to a specific theme, genre, or technique (e.g., narrative continuation, point-of-view change, material symbolism, or the “What if...” scenario). Texts are free in length and style, though shaped by the constraints of the given activity. SCRICREA is structured into eight sub-corpora, developed across six academic years (2019–2024). The total size exceeds one million words.

TABLE 1 Overview of lessons and prompts

Lessons	Topic/Prompts
1	Continuation of the story (The Young Crab - “il giovane gambero”)
2	Inventing a figure made of material (e.g., glass, wood) (James of Crystal - “Giacomo di Cristallo”)
3	Imaginary interview, diary, or autobiography of a protagonist (The Well of Cascina Piana)
4	Change of narrative perspective or interpolation (The Beautiful Stranger)
5	Music-inspired or object-protagonist narrative (Way on Clouds)
6	Genre change or “What if...” scenario (The Man Who Stole the Colosseum)
7	Verb substitution and text alteration (The King Who Had to Die)
8	Song-based narrative or free writing

TABLE 2 Corpus size by academic year and sub-corpus

Years	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7	Unit 8	Total
2019	13,951	34,995	36,664	34,400	33,261	25,101	19,600	26,451	224,423
2020	17,584	33,853	38,404	37,250	36,893	28,681	37,878	29,451	259,994
2021	19,018	37,399	40,765	41,853	42,008	34,416	23,121	39,965	278,545
2022	8,485	20,524	19,966	18,302	20,244	17,396	17,390	16,307	138,614
2023	6,974	10,833	10,415	11,507	11,489	5,552	10,308	8,411	75,489
2024	7,255	9,716	10,878	8,693	7,562	8,934	9,470	9,248	71,756
Total	73,267	147,320	157,092	152,005	151,457	120,080	117,767	129,833	1,048,821

Compared to larger reference corpora such as the Perugia Corpus [20] or the Cambridge Learner Corpus, SCRICREA is smaller in scale but unique in its exclusive focus on creative writing. While reference corpora aim at representativeness across genres, SCRICREA privileges depth in the pedagogical context, documenting creative processes under structured classroom conditions.

3. ANNOTATION AND ENRICHMENT

Texts in SCRICREA are systematically categorized by theme, genre, and task. Metadata includes learner age, gender, academic year, and task type. Annotation is multilayered: grammatical (POS-tagging with TreeTagger), lexical (n-gram frequency), and stylistic features (genre, creativity techniques). Following practices from other corpora [20],[21], annotation is both linguistic and structural, allowing for cross-linguistic comparisons and automatic evaluation.

4. APPLICATIONS AND PEDAGOGICAL RELEVANCE

SCRICREA serves as a dual-purpose resource: both as a research corpus for interlanguage analysis and as a pedagogical tool for language teaching. In line with research on learner corpora [3], [4], [22], SCRICREA allows for data-driven learning, raising learner awareness of linguistic forms and genres. Pedagogically, the corpus offers a wide range of applications that can enrich both language teaching and learning practices. One key use lies in the design of creative writing tasks that encourage learners to balance linguistic accuracy with imaginative expression. By drawing on authentic examples and linguistic patterns

within the corpus, teachers can guide students to explore language creatively while maintaining grammatical and stylistic precision [7], [6].

Furthermore, the corpus can serve as a powerful tool for enhancing genre awareness through contrastive analysis. By comparing and contrasting different narrative techniques across genres, learners can develop a deeper understanding of stylistic variation and discourse conventions, which in turn strengthens their ability to adapt their writing to diverse communicative contexts [22]. In addition, this corpus plays an important role in teacher education, particularly in cultivating familiarity with corpus-based methodologies. Training teachers to use corpora effectively enables them to integrate authentic language data into their lessons, promoting a more evidence-based and contextually grounded approach to language instruction [3]. Serving as a model for interdisciplinary approaches that combine corpus linguistics, second language acquisition, and creativity studies. Beyond classroom-based activities, SCRICREA has implications for corpus-informed syllabus design [4], curriculum innovation, and even automated feedback for learners. Incorporating SCRICREA data into digital writing platforms could allow teachers to provide individualized, data-driven feedback. Moreover, the corpus supports comparative analysis with other learner corpora, enabling cross-linguistic pedagogical insights.

Moreover, SCRICREA can be incorporated into digital learning environments and extended to secondary education, where it functions both as a source of authentic learner Italian and as a model for creative expression. Its multi-genre nature facilitates critical thinking and problem-solving through narrative experimentation, aligning with contemporary pedagogical frameworks that emphasize learner autonomy, creativity, and multimodality.

TABLE 3 Pedagogical applications of SCRICREA

Application Type	Corpus Feature Used	Pedagogical Outcome
Creative writing prompts	Sub-corpora by theme/genre	Magnify creativity & linguistic accuracy
Genre transformations	Genre-specific annotation	Genre awareness & flexibility
Music-inspired writing	Song-based sub-corpus	Emotional expression & multimodality
Error/contrastive analysis	Annotated learner output	Language awareness & peer learning

The SCRICREA corpus not only expands the methodological repertoire of learner corpus research but also presents several challenges. One limitation is its linguistic scope: as it focuses on Italian L2 learners with Greek as their L1, its representativeness is restricted. However, this narrow scope allows for depth of analysis and valuable insights into cross-linguistic influence. Another challenge is sustainability: corpus expansion depends on continuous student participation and institutional support. Technical challenges include anonymization of personal data, copyright issues related to prompts or music-inspired tasks, and the need for long-term digital preservation. Despite these limitations, SCRICREA contributes to the growing field of creative writing corpora, highlighting the importance of creativity in second language acquisition. Its interdisciplinary approach bridges applied linguistics, pedagogy, and creativity studies, paving the way for comparative initiatives in other languages.

5. EVALUATION AND FUTURE WORK

The SCRICREA corpus currently includes more than one million tokens and continues to grow annually. Its strengths lie in its interdisciplinary design, authenticity of learner production, and adaptability for multiple forms of analysis. Future work includes expanding access via online interfaces, refining annotation with NLP tools, and incorporating rubric-based evaluation for writing quality [21]. Long-term goals involve increasing its integration into curricula and extending its multilingual comparative potential. Future research will also explore how learners engage with the corpus as co-researchers, developing critical data literacy skills. The integration of SCRICREA into teacher training programs could further enhance its impact, equipping educators with tools to exploit learner data for pedagogical innovation.

6. CONCLUSION AND IMPLICATIONS

SCRICREA represents an innovative contribution to learner corpus research by merging corpus linguistics and creative writing [1], [2]. Its systematic structure, interdisciplinary scope, and pedagogical potential make it a valuable resource for both researchers and educators. By encouraging creativity alongside linguistic competence, SCRICREA highlights the productive intersection of language learning and artistic expression [6], [15].

SCRICREA also stands as a pioneering contribution to learner corpus research by integrating creative writing into second language pedagogy [23]. Its interdisciplinary design connects linguistic analysis with expressive freedom, offering a rich

empirical basis for exploring interlanguage phenomena, lexical development, and narrative competence [24], [8]. By estimulante grammatical accuracy, intercultural awareness, and learner agency [5], [14], [18]. The corpus not only supports data-driven instruction and curriculum innovation but also reimagines language education as a space for personal exploration, critical reflection, and authentic communication [7], [13], [16]. As such, SCRICREA stands as a replicable model for future initiatives that seek to combine corpus linguistics, creativity studies, and transformative pedagogy [11], [17].

From an institutional and technological perspective, the corpus points toward the growing importance of digital literacy and data-informed pedagogy in language education [3], [4]. Its integration into teacher training and curriculum design models how corpus-based methodologies can be harnessed to promote reflective teaching and evidence-based innovation [22]. The possibility of embedding SCRICREA data into digital platforms for automated or semi-automated feedback further illustrates the alignment between computational tools and creative pedagogy [21]. This convergence holds promise for the development of adaptive learning environments where learners and teachers become co-researchers in the process of linguistic discovery [23]. Finally, the implications of SCRICREA extend beyond Italian language education. As a replicable model, it demonstrates the viability of integrating creative writing corpora into other linguistic and cultural contexts [19]. Its interdisciplinary framework invites collaboration across applied linguistics, digital humanities, and education, encouraging the creation of learner corpora that capture the human, expressive dimension of language use [5], [15]. In doing so, SCRICREA not only advances research and pedagogy but also reaffirms the centrality of creativity in the cultivation of linguistic competence, intercultural empathy, and personal growth [14], [18].

ACKNOWLEDGEMENTS

Special thanks to all our students of the Department of Italian Language and Literature.

REFERENCES

- [1] S. Granger, "Computer Learner Corpus Research: Current Status and Future Prospects," BRILL eBooks, pp. 123–145, Jan. 2004, doi: https://doi.org/10.1163/9789004333772_008.
- [2] T. McEnery and A. Hardie, *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press, 2012.
- [3] F. Meunier, "Learner corpora and pedagogical applications," in *The Routledge handbook of language learning and technology*, L. Murray and F. Farr, Eds. London: Routledge, 2016, pp. 343–356.
- [4] U. Römer, "Corpus Research Applications in Second Language Teaching," *Annual Review of Applied Linguistics*, vol. 31, pp. 205–225, Mar. 2011, doi: <https://doi.org/10.1017/s0267190511000055>.
- [5] A.E.R. Ziad, "Creative writing and how it can support second language learners," Student thesis, Malmö Universitet, 2022. Available: <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1676349&dswid=6480>
- [6] T. B. Tin, "Language Creativity and Co-emergence of Form and Meaning in Creative Writing Tasks," *Applied Linguistics*, vol. 32, no. 2, pp. 215–235, Dec. 2010, doi: <https://doi.org/10.1093/applin/amq050>.
- [7] K. Arshavskaya, "Creative writing assignments in a second language course: A way to engage less motivated students," *Insight: A Journal of Scholarly Teaching*, vol. 10, pp. 68–78, 2015.
- [8] K. Florou, and I. Tyrou, "La particella 'ne' nella scrittura creativa in italiano LS, Indagine su un corpus di apprendenti greci," *Umanistica Digitale*, vol. 18, pp. 87-101, 2024, doi: <http://doi.org/10.6092/issn.2532-8816/19939>
- [9] Sylviane Granger and Gaëtanelle Gilquin, *Twenty years of learner corpus research : looking back, moving ahead*. Louvain-La-Neuve: Ucl, Presses Universitaires De Louvain, Cop, 2013.
- [10] B. L. Bruyn and M. Paquot, "Learner Corpus Research and Second Language Acquisition: an attempt at bridging the gap," Cambridge University Press eBooks, pp. 1–9, Dec. 2020, doi: <https://doi.org/10.1017/9781108674577.002>.
- [11] S. Mansoor, *Creative writing in language learning: A multidimensional approach*, Oxford: Oxford University Press, 2014.
- [12] Sirvy Mega Melvita, "Implementing Unfinished Story to Build Students' Creative Writing," *Journal of Educational Study*, vol. 3, no. 2, pp. 101–107, Dec. 2023, doi: <https://doi.org/10.36663/joes.v3i2.436>.
- [13] L. Nasir, S. Naqvi, and S. Bhamani, "Enhancing students' creative writing skills: An action research project," *Acta Didactica Napocensia*, vol. 6, no. 2, pp. 27–32, 2013.
- [14] P. Venuta, "Writing the wrong in the ELA classroom: The role of performance through creative writing," *LEARNing Landscapes*, vol. 13, no. 1, pp. 237–248, 2020.
- [15] D. I. Hanauer, "Chapter 1. Appreciating the beauty of second language poetry writing," *Linguistic approaches to literature*, pp. 11–22, Jun. 2014, doi: <https://doi.org/10.1075/lal.19.01han>.
- [16] L. Giacomini, "Using 'Storybird' in young learners' creative writing class," *English Teaching Forum*, pp. 35–37, 2015.
- [17] Tyrou, "Suggestions and prompts of foreign language activities in expressive writing," *International Journal of Education*, vol. 12, no. 2, 2024.
- [18] Matea Butković and E. Vidović, "Fostering Intercultural Dialogue through Literature and Creative Writing: Examples of Successful Classroom Practices," pp. 229–248, Dec. 2021, doi: <https://doi.org/10.26493/978-961-7055-36-8.229-248>.
- [19] H. Zheng, "Integrating Creative Writing and Anthropology in General Education: A Better Approach to Cultural Understanding and Critical Thinking," *Iris Journal of Educational Research*, vol. 4, no. 3, Nov. 2024, doi: <https://doi.org/10.33552/ijer.2024.04.000588>.

- [20] S. Spina, "Il Perugia Corpus: una risorsa di riferimento per l'italiano," Proceedings of CLiC-it 2014 & EVALITA 2014, Pisa: Pisa University Press, 2014, pp. 354–359.
- [21] Song, K. T. Lim, and J. Park, "Enriching the Korean learner corpus with multi-reference annotations," arXiv preprint, 2025. [Online]. Available: <https://arxiv.org/abs/2505.00261>
- [22] S. Staples, N. Conrad, and A. Dang, "Building language and genre awareness through learner corpus data," *International Journal of Learner Corpus Research*, vol. 10, no. 1, pp. 146-182, 2024. doi: <https://doi.org/10.1075/ijlcr.00043.sta>
- [23] Tyrou and K. Florou, "Learner corpus of creative writing: An interdisciplinary challenge," in *AIUCD 2025 Conference Proceedings*, Verona, Italy, June 11–13, 2025.
- [24] Aijmer, "Modality in advanced Swedish learners' written interlanguage," *Language learning and language teaching*, pp. 55–76, Jan. 2002, doi: <https://doi.org/10.1075/lllt.6.07aij>.