



SOCIO-POLITICAL NEOLOGISMS IN SKETCH ENGINE CORPORA: A CROSS-LINGUISTIC FRENCH–SPANISH–SLOVAK COMPARATIVE STUDY

SOCIÁLNO-POLITICKÉ NEOLOGIZMY V KORPUSOCH SKETCH ENGINE: MEDZIJAZYKOVÁ FRANCÚZSKO-ŠPANIELSKO-SLOVENSKÁ KOMPARATÍVNA ŠTÚDIA

LUCIA RÁČKOVÁ¹ – EVA REICHWALDEROVÁ²

¹Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici, Tajovského 40, 97401 Banská Bystrica, lucia.rackova@yahoo.com

²Filozofická fakulta, Univerzita Mateja Bela v Banskej Bystrici, Tajovského 40, 97401 Banská Bystrica, eva.reichwalderova@umb.sk

Abstract

This paper examines a selection of contemporary socio-political neologisms in French, with the primary objective of identifying their possible neologistic equivalents in Spanish and Slovak. To achieve this aim, we analyse data drawn from three comparable corpora of present-day language available in *Sketch Engine: frTenTen23*, *esTenTen23*, and *skTenTen23*. These linguistic resources provide valuable evidence about the frequency, contextual usage, and collocational behaviour of recently coined lexical units. By contrasting the corpus data, we aim to observe how neologisms originating in French can be adapted, translated, or calqued in Spanish and Slovak. We hypothesize that the identification of neologistic equivalents will be easier in Spanish than in Slovak due to their closer morphological and lexical proximity to French. Furthermore, the study explores productive word-formation suffixes and their cross-linguistic correspondences. The findings are expected to contribute to research on lexical innovation and have practical applications in French–Spanish, French–Slovak, and Spanish–Slovak translation studies.

Keywords

suffixation, noun, contrastive neology, *TenTen* family corpora

Introduction

Among derivational mechanisms, suffixation meets speakers' need to integrate emerging vocabulary into their language, particularly within the political sphere. This simple yet highly productive process plays a crucial role in the formation of newly coined nouns. As language functions as an essential medium for communicating new phenomena to the public, it gains particular prominence in times of crisis and socio-political upheaval (Kwofie, 1979; Yegana, 2018; Kacprzak, 2019). Each language, and consequently each linguistic system, is organized to foster lexical creativity through a wide range of morphological and semantic patterns (Pruvost – Sablayrolles, 2003). The lexicon is the most dynamic and evolving component of any language (Timko, 2015, 71).

The aim of this paper is to identify Spanish and Slovak equivalents of a selective sample of French socio-political neologisms formed through suffixation. Two criteria are applied in selecting equivalents. First, since the study focuses on neologistic transfer, the equivalents themselves should also display neologistic status; that is, they should be new lexemes not yet recorded in standard dictionaries. Second, their use in contemporary language should be attested by their occurrence in the *Sketch Engine TenTen* corpora. The comparative approach adopted in this study highlights the dynamics of suffixal derivation and offers valuable insights into how French, Spanish, and Slovak function from a contrastive perspective. This interdisciplinary trilingual research lies at the intersection of Natural Language Processing, Political Discourse Analysis, Word Formation, Contrastive Neology, Language Contact, and Translation Studies.

1 Theoretical Framework of the Research

In presenting our research, we will begin with the theoretical background for the analysis of neologistic suffixes in contemporary French and subsequently in Spanish and Slovak. We will first outline key approaches to suffixes as a mechanism of lexical innovation, with a particular emphasis on nominal neologisms and their productivity,



while also presenting the definition of neologisms used in this study. The framework further addresses the issues of linguistic and cultural untranslatability based on specific *Sketch Engine TenTen* corpora.

1.1 Suffixation in Lexical Innovation

One of the earliest variationist studies based on a corpus devoted to French affixation is that of Grabar et al. (2006), who analysed a 25-million-word corpus of journalistic texts from across the French-speaking world. Huyghe – Lombard (2022) note that the suffix *-age* is regarded as one of the most productive nominalizing suffixes in French. Plag et al. (2018) examined deverbal neologisms ending in *-ment* in English, while Dal et al. (2018) compared French nouns ending in *-age* and *-ment* as used in web-sourced texts. According to Dubois (1962) and Lüdtke (1978), the suffix *-age* tends to attach to transitive bases, whereas *-ment* more commonly occurs with intransitive ones. Similarly, the suffix *-isme* is also highly productive, as reflected in its long-standing use, with derivatives already attested in the first edition of the *Dictionnaire de l'Académie française* (1694) (Holeš, 2024, 5).

Following Sablayrolles (as cited in Mudrochová, 2021), neologisms are defined by their absence from reference dictionaries (*Trésor de la langue française informatisé*, *Dictionnaire de l'Académie française*, *Le Robert Dico en ligne* for French; *DLE – Diccionario de la lengua española* of the RAE – *Real Academia Española* for Spanish; and the *Slovníkový portál JÚLŠ* for Slovak). Neologisms tied to new or culturally bound concepts are deeply embedded in each society's linguistic context, making them difficult to render across languages. According to Catford's (1965) framework of linguistic untranslatability, these forms can be seen as expressions resisting exact translation.

1.2 Language Resources: Sketch Engine TenTen Corpora

Neologisms are identified in corpora available in *Sketch Engine*, a Czech corpus management tool developed in 2004 at Masaryk University in Brno by Adam Kilgarriff and Pavel Rychlý. In *Sketch Engine*, corpus data are automatically lemmatized and parts of speech tagged. More specifically, we focus on contemporary texts: *frTenTen23* (27 billion tokens), *esTenTen23* (33 billion tokens), and *skTenTen23* (1 billion tokens). All three corpora cover the widest possible range of genres, topics, text types, and web sources. They are suitable for both general and specialized language research. Although the corpus names suggest that the texts date from 2023, some of them were in fact collected earlier. In the *TenTen* corpora, texts are separated into individual positions (tokens) after being collected from the Internet by the Spiderling tool, a 'web spider' designed for linguistic purposes, and cleaned by jusText, which removes undesirable content.

2 Methodology and Research Hypothesis

Drawing on data from *Sketch Engine* (Chalupníková – Volková, 2013; Thomas et al., 2005; Evans, 2021), we hypothesize that while most French neologisms tend to have Spanish equivalents, their transfer into Slovak is less straightforward due to weaker linguistic contact and morphological dissimilarity. To test this, we use the *Concordance* and *Good Dictionary Example (GDEX)* functions of *Sketch Engine*, supported by a frequency analysis of suffixes and neologistic equivalents in both target languages. The study, based on the *frTenTen23*, *esTenTen23*, and *skTenTen23* corpora, also assesses *Sketch Engine*'s potential for comparative research on lexical innovation.

3 Comparative Analysis of Neologisms

As mentioned above, this study focuses on French suffixed neologisms that mirror not only the French political landscape, such as *macronisme* or *miterrandisme*, but also international and Latin American contexts, as in *guévarisme* (from *Che Guevara*) or *lulisme* (from *Lula da Silva*). The research is based on a corpus of French socio-political neologisms identified in linguistic studies, media discourse, the *Néoveille* platform, and *Sketch Engine* during searches for other new lexical items. We would just like to add that the *Néoveille* platform is the result of a research project funded by the IDEX of the COMUE Sorbonne Paris Cité, led by Professor Sablayrolles' team (Cartier, 2019), which aims to track the life cycle of neologistic lexical units in a dynamic contemporary corpus.

Using *Sketch Engine*'s concordance tool, we examined the meanings of these neologisms and their equivalents in authentic contexts. However, in some cases, French neologisms are rendered by non-neologistic counterparts, with or without derivational suffixes, or by multi-word expressions. For example, *macronerie*, which we translate into Spanish as *política macronista* and into Slovak as *Macronova politika*.

A cross-linguistic comparison of parallel French, Spanish, and Slovak suffixal neologisms is presented below. To ensure the objectivity of the analysis, frequencies are expressed as relative frequencies in scientific notation. This



approach was preferred over absolute frequencies, as the corpora differ in size. However, this is problematic because *Sketch Engine* does not always rely on the same method for expressing relative frequency. Therefore, all data were normalized and are reported in scientific notation (e.g., 1e-7%), corresponding to 1×10^{-7} .

A) French Suffix -ISATION

French: extrémisation (6.9e-7%), kosovarisation (0.15e-7%), trumpisation (8.9e-7%)

Spanish: extremización (0.24e-7%), kosovarización (0.03e-7%), trumpización (1.1e-7%)

Slovak: extremizácia (9.2e-7%), kosovizácia (0.92e-7%), trumpizácia (1.8e-7%)

B) French Suffix -ISME

French: bolivarianisme (1.4e-7%), islamo-gauchisme (150e-7%), poutinisme (21e-7%), trumpisme (72e-7%)

Spanish: bolivarianismo (61e-7%), islamoizquierdismo (1e-7%), putinismo (14e-7%), trumpismo (130e-7%)

Slovak: bolivarizmus (0.92e-7%), islamoľavičiarstvo (0.92e-7%), putinizmus (40e-7%), trumpizmus (11e-7%)

C) French Suffix -ISTE

French: chaviste (93e-7%), enfermiste (1.4e-7%), trumpiste (43e-7%)

Spanish: chavista (1500e-7%), aislacionista (150e-7%), trumpista (80e-7%)

Slovak: chavista (0.92e-7%), izolacionista (3.7e-7%), trumpista (0.92e-7%).

In terms of realized productivity, the number of neologistic units is the same across all three languages for the suffixes listed above.

In French, the lexeme *trumpisation* enjoys particularly high frequency, being deeply rooted in French media discourse. Its occurrence in French is eight times higher than its equivalent in Spanish (*trumpización*) and almost five times higher than its Slovak equivalent (*trumpizácia*). Conversely, *trumpismo* is most frequent in Spanish, occurring 1.8 times more often than in French (*trumpisme*) and almost twelve times more often than in Slovak (*trumpizmus*). Likewise, *trumpista* appears nearly twice as often in Spanish as in French (*trumpiste*), and about 86 times more often than in Slovak (*trumpista*). Similarly, the French lexeme *islamo-gauchisme* occurs approximately 150 times more frequently than its Spanish counterpart (*islamoizquierdismo*) and 163 times more than in Slovak (*islamoľavičiarstvo*).

Given the Latin American context, it is unsurprising that the neologism *bolivarianismo* is most frequent in Spanish, 43 times more frequent than in French (*bolivarianisme*) and 66 times more than in Slovak (*bolivarizmus*). Spanish also shows a markedly higher frequency for *chavista*, with values 16 times greater than in French (*chaviste*) and about 1,600 times higher than in Slovak (*chavista*). Interestingly, the same pattern is observed for *aislacionista*, which occurs 107 times more often in Spanish than in French (*enfermiste*) and 40 times more than in Slovak (*izolacionista*).

The Slovak lexeme *extremizácia* represents a relatively frequent neologism, with a frequency 1.3 times higher than in French (*extrémisation*), while its Spanish equivalent (*extremización*) is 38 times less frequent. Similarly, *kosovizácia* occurs 6 times more often in Slovak than in French (*kosovarisation*) and 30 times more than in Spanish (*kosovarización*), although overall frequencies remain low. The Slovak form *putinizmus* shows nearly double the frequency of its French equivalent (*poutinisme*) and triple that of the Spanish one (*putinismo*).

Out of the twenty-five French neologisms examined, twenty-one corresponding neologistic equivalents (84%) were identified in Spanish. For each French derivational suffix, a single Spanish suffixal counterpart was found: -ade/-ada, -cide/-cidio, -ie/-ía, -isation/-ización, -isme/-ismo, -iste/-ista, and -ité/-idad.

For the same French neologisms, 10 corresponding neologistic equivalents, i.e. 40%, were identified in Slovak. To preserve the international character of the new lexical items, the French suffix *-isation* is consistently rendered in Slovak as *-izácia*, and *-iste* as *-ista*. In the case of *-isme*, the most frequent Slovak equivalent is *-izmus*, although the Slavic suffix *-stvo* also occurs as a possible variant.

4 Discussion of Results

For Spanish, the lexical unit *aislacionista* is cited in dictionaries only as an adjective; however, corpus evidence shows that it also occurs as a noun. Likewise, although the *Diccionario de la Lengua Española (RAE)* defines



lulismo merely as a “philosophical and mystical trend based on the thinking of Raimundo Lulio,” the *esTenTen23* corpus attests numerous examples referring to Lula da Silva, such as “una década del lulismo en el poder.” In contrast, no Slovak equivalent for *lulismo* was identified, as the term does not appear in the Slovak cultural and social context and therefore not even in the online environment. It should be noted that the *TenTen* corpora are web-based, which is why they reflect the vitality of contemporary language. In this context, Chovancová (2012) highlights the dynamics of online language and the graphic adaptations characteristic of interactions in digital environments. Further observations on emerging scriptural practices in chat rooms and other forms of electronic discourse, including humorous or ironic uses, which are also characteristic of neologization, are discussed in Lazar (2017).

Our initial hypothesis was confirmed. We found that in Spanish, up to 84 percent of French neologisms have neologistic equivalents. In Slovak, the proportion was roughly half, at only 40 percent. These findings would, of course, need to be verified on a larger trilingual research sample.

Regarding the evaluation and usefulness of *Sketch Engine*, in theory, many of its functions can be effectively applied to our research. However, in practice, the results obtained are often inconsistent or only partially relevant. For instance, when searching for suffixal neologisms using the asterisk operator (Kleene star) to identify words ending with specific suffixes, the output of neologistic units proved to have marginal value. At present, one of the main limitations lies in the fact that the semantic dimension of words cannot be automatically processed. Moreover, absolute frequencies are not objectively comparable due to the differing sizes of the corpora, while relative frequencies are often difficult to interpret, given the considerable variation in word occurrences.

Conclusion

Although *Sketch Engine* offers a wide range of valuable functionalities, we are still far from achieving fully automatic detection of neologisms through corpus tools. Due to the limited availability of parallel corpora, this investigation had to be carried out separately in three comparable datasets. Furthermore, the Slovak corpus *skTenTen23* is insufficient, as three-quarters of its content consists of non-specialized texts, highlighting the limitations inherent in corpora for smaller languages.

The results indicate that linguistic, geographical and socio-political proximity between French and Spanish led to more than twice as many neologistic equivalents as between French and Slovak. French and Spanish share a number of common suffixes, reflecting their shared Latin origin and close linguistic proximity. Slovak, by contrast, displays greater morphological variability, combining Latin-derived suffixes with its own native Slavic endings. While clear suffixal parallels can be observed between French and Spanish, Slovak tends to rely more on processes of domestication and descriptive translation. In addition, Slovak has greater variability in suffixes.

Consequently, the translation of new lexical units from French into Slovak continues to represent a major linguistic and cultural challenge. The translatability of neologisms likely depends not only on a language’s morphological structure but also on the degree of linguistic and geographical contact, which makes their transfer into Slovak less straightforward. Furthermore, the Slovak corpus *skTenTen23* is limited in its suitability for in-depth research, as three-quarters of its content consists of non-specific texts. This highlights the limitations inherent in corpus resources for smaller languages.

Acknowledgements: Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00417. Within the framework of more efficient governance and strengthening RDI funding – Investment. Excellent science.



Funded by the
European Union
NextGenerationEU

References

- ACADÉMIE FRANÇAISE, [n. d.]. *Dictionnaire de l'Académie française. 9e édition* [online]. Paris: Académie française. Available online: <https://www.dictionnaire-academie.fr/> [cit. 2025-10-18].
- CARTIER, Emmanuel. 2019. *Néoveille, plateforme de repérage et de suivi des néologismes en corpus dynamique. Neologica*, Vol. 13, pp. 23–54.
- CATFORD, John Cunnison. 1965. *A linguistic theory of translation*. Oxford: Oxford University Press, 103 pp. ISBN 978-0194370189.



- CHALUPNÍKOVÁ, Daniela – VOLKOVÁ, Nikol. 2013. *Dotazy v CQL (pro Sketch Engine), stručný přehled CQL s důrazem na vyhledávání v českých korpusech*. In *PLIN021: Sémantická analýza v praxi*.
- CHOVANCOVÁ, Katarína. 2012. *La communication médiée*. Časopis Ústavu filologických štúdií Pedagogickej fakulty UK v Bratislave, Vol. 22, No. 1–2, pp. 23–27.
- DAL, Georgette – HATHOUT, Nabil – LIGNON, Stéphanie – NAMER, Fiammetta – TANGUY, Ludovic. 2018. *Toile versus dictionnaires: Les nominalisations du français en -age et en -ment*. In NEVEU, F. – HARMEGNIES, R. – HRIBA, L. – PRÉVOST, S. (eds.), *Congrès mondial de linguistique française*. Paris: Institut de Linguistique Française, pp. 1–12.
- DUBOIS, Jean. 1962. *Étude sur la dérivation suffixale en français moderne et contemporain*. Paris: Librairie Larousse, 118 pp.
- esTenTen23 Corpus. 2023. In *Sketch Engine. Lexical Computing Ltd.* [online]. Available on: <https://www.sketchengine.eu> [cit. 2025-10-18].
- EVANS, Mel. 2021. Kilgariff, Adam, founder; Pavel Rychlý, co-founder; and Miloš Jakubíček, CEO. Sketch Engine. In *Renaissance and Reformation*, Vol. 44, No. 4, pp. 217–218.
- frTenTen23 Corpus. 2023. In *Sketch Engine. Lexical Computing Ltd.* [online]. Available on: <https://www.sketchengine.eu> [cit. 2025-10-18].
- GRABAR, Natalia – TRIBOUT, Delphine – DAL, Georgette – FRADIN, Bernard – HATHOUT, Nabil – LIGNON, Stéphanie – NAMER, Fiammetta – PLANCO, Clément – YVON, François – ZWEIGENBAUM, Pierre. 2006. *Productivité quantitative des suffixations par -ité et -able dans un corpus journalistique moderne*. In *Actes de la 13e conférence sur le Traitement Automatique des Langues Naturelles*, pp. 167–177.
- HOLEŠ, Jan. 2024. *Mots dérivés de noms propres de personnes dans la neuvième édition du Dictionnaire de l'Académie française. Le cas des mots formés à l'aide du suffixe -isme*. In *L'Analisi Linguistica e Letteraria*, Vol. 32, No. 3, pp. 5–20.
- HUYGHE, Richard – LOMBARD, Alizée. 2022. *Les néologismes en -age en français contemporain: Héritage verbal et polysémie*. In *Journal of French Language Studies*, Vol. 32, pp. 25–47. Cambridge: Cambridge University Press.
- KACPRZAK, Alicja. 2019. *Dans le monde de la macronie ou quelques notes sur les néologismes éponymiques du domaine politique en français contemporain*. In *Voces dialogantes. Estudios en homenaje al profesor Waczesław Nowikow*, pp. 1–10. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- KWOFIE, Emmanuel N. 1979. *Une étude sur le suffixe -ation en français*. In: *Revue Romane*, Vol. 14, No. 2, pp. 123–140.
- LAZAR, Jan. 2017. *À propos des pratiques scripturales dans l'espace virtuel : entre Facebook et Twitter*. Ostrava: Faculty of Arts, University of Ostrava.
- LE ROBERT, [n. d.]. *Le Robert en ligne* [online]. Paris: Le Robert. Available online: <https://dictionnaire.lerobert.com/> [cit. 2025-10-18].
- LÜDTKE, Jens. 1978. *Prädikative Nominalisierungen mit Suffixen im Französischen, Katalanischen und Spanischen*. Berlin, Boston: Max Niemeyer Verlag, pp. 87–219. <https://doi.org/10.1515/9783111662879-004>
- MUDROCHOVÁ, Radka. 2021. *Jean-François Sablayrolles: Comprendre la néologie. Conceptions, analyses, emplois*. In *Časopis pro moderní filologii*, Vol. 103, No. 2, pp. 295–302.
- PLAG, Ingo – ANDREOU, Marios – KAWALETZ, Lea. 2018. *A frame-semantic approach to polysemy in affixation*. In BONAMI, O. – BOYÉ, G. – DAL, G. – GIRAUDO, H. – NAMER, F. (eds.) *The lexeme in descriptive and theoretical morphology*. Berlin: Language Science Press, pp. 467–486.
- PRUVOST, Jean – SABLAYROLLES, Jean-François. 2003. *Les néologismes*. Paris: Presses Universitaires de France, 128 pp.
- REAL ACADEMIA ESPAÑOLA, 2023. *Diccionario de la lengua española (DLE)* [online]. Madrid: RAE. Available online: <https://dle.rae.es/> [cit. 2025-10-18].



SKETCH ENGINE, [n. d.]. *TenTen corpora* [online]. Brno: Sketch Engine. Available online: <https://www.sketchengine.eu/documentation/tenten-corpora/> [cit. 2025-10-18].

skTenTen23 Corpus (2023). In *Sketch Engine. Lexical Computing Ltd.* [online]. Available on: <https://www.sketchengine.eu> [cit. 2025-10-18].

SLOVNÍKY JAZYKOVEDNÉHO ÚSTAVU EUDOVÍTA ŠTÚRA SAV, [n. d.]. [online]. Bratislava: *Jazykovedný ústav Ľudovíta Štúra SAV*. Available online: <https://slovník.juls.savba.sk/> [cit. 2025-10-18].

THOMAS, James Edward – KILGARIFF, Adam – SMITH, Simon – MARCOWITZ, Fredrik. 2015. Corpora and language learning with the Sketch Engine and SKELL. In *Revue Française de Linguistique Appliquée*, Vol. 20, No. 1, pp. 37–48.

TIMKO, Andrej. 2015. Anglicisms, pseudo-anglicisms and franglais in language of French gastronomy. In *Acta Linguistica*, No. 10, pp. 69–78. Banská Bystrica: Univerzita Mateja Bela v Banskej Bystrici, Ekonomická fakulta, Katedra odbornej jazykovej komunikácie.

TRÉSOR DE LA LANGUE FRANÇAISE INFORMATISÉ, 2012. [online]. Nancy: ATILF – CNRS & Université de Lorraine. Available online: <http://www.cnrtl.fr/definition/> [cit. 2025-10-18].

YEGANA, Orujova. 2018. Les particularités du discours politique français. In *Revista académica liLETRAd*, Vol. 4, pp. 277–284. Baku: Azerbaijan University of Languages.