

Developing a Digital, Usage-Oriented, Multilingual Terminology Resource on Transgender Vocabulary: Methodological Challenges, Limitations and Possible Solutions*

Anna Jamka^{1,*†}, Silvia Bonaverò^{2,3,†}, Océane Foubert^{4,†}, Miguel Sánchez Ibáñez^{5,†}, Giovanni Tallarico^{2,†} and Aleksandra Tomaszewska^{6,†}

¹ University of Warsaw, Krakowskie Przedmieście 26/28, 00-927 Warsaw, Poland

² University of Verona, Lungadige Porta Vittoria 41, 37129 Verona, Italy

³ University of Paris Cité, 85 boulevard Saint-Germain, 75006 Paris, France I

⁴ University of Picardy Jules Verne, Chemin du Thil, 80025 Amiens, France

⁵ Facultad de Filosofía y Letras, Universidad de Valladolid, Plaza del Campus Universitario s/n. 47400, Valladolid, Spain

⁶ Institute of Computer Science, Polish Academy of Sciences, ul. Jana Kazimierza 5, 01-248 Warsaw, Poland

Abstract

This paper presents the development of a multilingual, usage-oriented terminology resource on trans language. It comprises an annotated glossary documenting key trans-related terms across English, Polish, Spanish, French, and Italian. Adopting a hybrid approach that integrates structured terminology with corpus-documented variation, the resource serves both general audiences, seeking trans-inclusive language guidance, and researchers studying lexical innovation and trans linguistics. We outline conceptual, methodological, and technical challenges encountered during glossary development, like selection criteria limitations, concept definition complexities, corpus adequacy for studying rapidly-evolving vocabulary, and terminological difficulties inherent to community-driven, non-standardized lexis. The project intends to fill gaps in existing resources by providing methodological transparency as well as cross-linguistic documentation of trans terminology, while highlighting future research directions in dissident community-driven linguistic innovation.

Keywords

linguistic innovation, trans linguistics, glossary, multilingual terminology

1. Introduction

This paper outlines the process of developing a digital, usage-oriented, multilingual terminology resource on trans language, undertaken by a research group of linguists within COST Action CA22126: European Network on Lexical Innovation (ENEOLI) [1]. The resource, hereinafter referred to as the “tool”, is planned to consist of an annotated glossary, documenting key terms related to trans experiences, and an outline of grammatical and communicative strategies used by and to refer to trans individuals. The tool is intended to provide both a practical guide for trans inclusive language, and a methodological reference for scholarly research.

In the paper, we examine the conceptual, methodological, and technical challenges and limitations that arose during the first phase of the project—developing a multilingual glossary—as well as discuss the solutions to address them. The second part of the project, i.e., the grammatical and communicative strategies, are not discussed here.

*5th International Conference on “Multilingual digital terminology today. Design, representation formats and management systems” (MDTT) 2026, June 25-26, 2026, Zadar, Croatia.

†Corresponding author.

†These authors contributed equally.

✉ anna.jamka@uw.edu.pl (A. Jamka); silvia.bonaverò@univr.it (S. Bonaverò); oceane.foubert@u-picardie.fr (O. Foubert); miguel.sanchez@uva.es (M. Sánchez Ibáñez); giovanni.tallarico@univr.it (G. Tallarico); aleksandratomaszewska15@gmail.com (A. Tomaszewska)

ORCID 0000-0001-8201-6476 (A. Jamka); 0009-0002-7679-0670 (S. Bonaverò); 0000-0002-7015-5767 (O. Foubert); 0000-0001-5712-7553 (M. Sánchez Ibáñez); 0000-0002-0585-4473 (G. Tallarico); 0000-0001-6379-3034 (A. Tomaszewska).



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The paper is structured as follows. First, it introduces the resource under development and situates it among existing linguistic resources and scholarship on trans language and neology in the European varieties of the five languages studied: English, Polish, Spanish, French, and Italian. Next, it describes materials and methods used. The discussion section presents and justifies the conceptual and methodological decisions made throughout the project. Finally, the conclusion summarizes the main contributions and suggests lines for future research.

2. The tool: rationale, objectives, target audience

The tool is designed as a multilingual, annotated glossary of key terms related to trans experiences, accompanied by a description of grammatical and communicative strategies used by trans people and to refer to trans people.

It has a two-fold purpose. First and foremost, it is designed to support the general public and language professionals (teachers, translators, drafters, journalists, etc.) in using trans-inclusive language. Research-wise, it aims to provide a reliable point of reference for linguists interested in the intersection of linguistic innovation and trans linguistics. It is also intended to be a valuable source of information for historians, sociologists, communications scholars, and any other kind of professionals interested in tracking the evolution of vocabulary related to trans experiences.

The project was motivated by two main factors. First, although a wide range of LGBTQ+ glossaries and guidelines already exist [2, 3, 4, 5, 6], most are developed outside academia and may therefore lack the methodological transparency required for further evidence-based research. Conversely, much of the existing scholarly work on trans-related vocabulary focuses on, for example, its linguistic innovativeness [7], specific contexts where such lexis is used, such as transgender health [8, 9], translation studies [10], or education [11].

These studies are predominantly monolingual and are not designed to function as general terminological resources. We think it is important to adopt a multilingual approach for both practical and epistemological reasons.

Trans identities and experiences are articulated through culturally specific linguistic frameworks that usually resist direct translation—a reality that reflects genuine differences in how gender diversity is conceptualized across communities rather than mere semantic variation [12]. English-language terms like "non-binary" or "genderqueer" carry particular cultural and historical baggage that may not align with how gender variance is understood in other linguistic contexts, where indigenous or locally developed concepts (such as "two-Spirit" in North America, "hijra" in India or "travesti" in Spain) articulate different relationships between identity, community, and social roles [13, 14]).

Beyond this theoretical consideration, our tool aims to address urgent practical needs: trans people navigate healthcare systems, legal frameworks, and social services that operate in multiple languages, and the absence of standardized, community-endorsed terminology in these contexts can lead to miscommunication, misgendering, and barriers to care [15]. A multilingual approach thus serves both to preserve the conceptual diversity of trans experiences globally and to ensure transparency and equitable access to representative language across linguistic communities [16].

Regarding the role played by terminology in the tool's development, its already mentioned two-fold purpose makes us consider carefully the way we manage specialized vocabulary. It is true that a merely classic terminological approach would offer systematic precision essential for translation, healthcare, and legal contexts, as well as ensure accurate cross-linguistic communication and reduce harm through standardized definitions [17]. However, rigid terminological frameworks risk institutionalizing language that remains fluid within trans communities, potentially replicating medical gatekeeping and erasing contestation over meaning [16]. It is our intention to take into account a socioterminological [18], corpus-based approach in order to better capture linguistic variation, community agency, and evolving usage [19], honoring self-determination over prescription. Yet this would jeopardize cross-linguistic comparability and practical precision that multilingual terminology provides to translators and service providers [20]. In consequence, we

intend to set a hybrid model integrating structured terminology with documented variation, in order to balance these competing needs.

This project was designed to fill the gaps identified.

3. Literature review

Existing glossaries of LGBTQ+ vocabulary and guidelines on trans inclusive communication are mostly community-based [4] or produced by institutions [2]. While they are valuable resources, their conceptual and methodological design is rarely made explicit, which limits their applicability for systematic linguistic research. This gap points to the need for a methodologically transparent resource such as the one proposed here.

This project is situated at the crossroads of trans linguistics and lexical innovation.

Trans linguistics [19] focuses on the study of the linguistic practices of gender-non-normative people, dealing with trans grammar, trans discourse, and trans voices. As a branch of linguistic studies, it challenges cisnormative assumptions in language systems, and examines agency in linguistic self-determination [21]. Rejecting pathologization [22], it treats voice and embodiment as material sites of gender production [23], while maintaining explicit political commitment to social justice [22]. Analyzing trans-related terminology could be thus understood as contributing to the way trans discourses are generated and disseminated.

At the same time, the project draws on scholarship on lexical innovation, which addresses the emergence, diffusion, and functioning of new linguistic forms. Adopting a broader perspective on lexical innovation, namely seeing it as a multidimensional phenomenon [24], allows us to situate trans terminology within wider processes of linguistic creativity. Trans vocabulary is also subject to ongoing changes to respond to dynamic shifts and developments, due to, on the one hand, the rapid outdateding of terminological frameworks in such a changing field, and, on the other, to the new coinages constantly emerging from community practice [23, 25]. This further connects this field to neology.

A survey of existing scholarship on trans language, in particular, trans vocabulary, in the languages of the project—English, Polish, Spanish, French, and Italian—reveals distinct methodologies and levels of development.

Most scholarly work dealing explicitly with English transgender terminology is situated within specialized contexts of LGBTQ+ health [8], social sciences research [26], or scholarly work in general [27]. Notably, we did not identify studies or terminological resources produced in academia that rely on corpus analyses.

Conversely, studies on trans vocabulary in Polish do not focus on specialized domains. Scholarly works on the topic deal with grammatical and communicative strategies, highlighting their innovativeness, rather than with terminological aspects of such lexis [7, 28, 29]. The most exhaustive terminological resource on trans experiences in Polish is *Słownik terminologii queerowej* (*Dictionary of queer terminology*) made available on zaimki.pl (English version available at: <https://en.pronouns.page/>), a platform developed by Rada Języka Neutralnego (the “Neutral Language Council” collective), which is a community-based, partially multilingual and digital resource in continuous development [4].

Scholarly work on trans linguistics in Spanish is often part of a larger effort to analyze and describe inclusive language [30]. One of the key themes under study in this field is morphological innovativeness of trans and non-binary speakers [31], most notably, the so-called inclusive (neo)morphemes *-x* and/or *-e* [32, 33]. Trans and non-binary Spanish is also often discussed in the context of translation studies [10, 34] and language teaching [35]. To the best of our knowledge, more terminology-oriented work on LGBTQ+ vocabulary in Spanish produced within academia concerns overviews of (scarce) existing resources [36] or more conceptual perspectives [37, 38]. Notably, there are also papers at the intersection of neology and queer linguistics, yet they are not limited to trans language [39]. There are also numerous community-based (i.e., [40]) and

institutional (i.e., [41]) glossaries and other resources on LGBTQ+ vocabulary, yet, more often than not, their methodological design is not made explicit.

Most of the work done on French focuses particularly on grammatical aspects. Linguistic research has been carried out on French inclusive grammar [42], non-binary language [43], and in the context of teaching French as a foreign language [44]. Regarding work on transgender terminology, a glossary has been published by the Canadian government [6].

Research on Italian trans vocabulary remains limited and has primarily focused on innovative linguistic forms, such as those using the schwa or the asterisk [45, 46]. New research also takes into account innovative AGEDs (Alternative grammatical Gender Encoding Devices), like diphthongs [47], even if it still remains a restraint phenomenon. However, these academic interests do not focus solely on trans vocabulary and rather stem from a broader debate on inclusive language. On the other hand, the majority of existing glossaries are community-based, as shown by the Modena Pride [48] and Torino Pride [49] glossaries, although not exclusively. Moreover, some university glossaries on inclusive language (i.e., [50, 51]) have partially addressed transgender terminology, too.

Taken together, the review of existing resources and scholarship on trans vocabulary across the languages studied demonstrates some limitations of the current landscape. As such, it indicates the need for a methodologically transparent and usage-oriented tool as the one developed in this project.

4. Material and methods

The initial set of 30 units (Appendix A), both single-word and multi-word expressions, selected for cross-linguistic examination was sourced from the Gender and Sexual Diversity Glossary [2] published by the Translation Bureau of the Government of Canada. Three criteria guided the inclusion in the study: (1) their relevance to gender identity rather than sexual orientation; (2) their role in understanding trans experiences; (3) their perceived degree of novelty.

Each unit was first treated as a concept. It was defined and classified into one of four available domains—social, legal, medicine, linguistics—by a single researcher.

Subsequently, each researcher was assigned a subset of units to analyze within their respective languages. The analyses were carried out using corpus methods, drawing on the Trends corpora [52] available in Sketch Engine [53]. To ensure methodological comparability across languages, we developed a common query protocol (Appendix B). On the basis of corpus evidence, each researcher was tasked with producing a structured description of the unit, comprising the following categories: part of speech, formal variants, corpus evidence (frequencies, collocations, usage patterns).

The entry schema for the final version of the tool is still under development. Given the tool's two-fold orientation (general and specialized audience), it needs to be carefully designed so as not to overwhelm a non-specialized user and, at the same time, be useful for linguistic research. The preliminary design of each entry comprises six core categories: concept definition, plain-language concept definition, main equivalent, variants, usage example, and usage labels (i.e. dated, recommended, not recommended). Information for a more specialized audience would be available via expandable fields such as part of speech, neologism type, comparative notes on usage, frequency data, and sociolinguistic observations such as discursive framing and other topics.

A two-level validation procedure was designed to ensure the reliability and relevance of the glossary. In the first stage, a language expert for each language will review the proposed equivalents, sources, PoS assignments, and the identification of the variants, neological type, etc. In the second stage, a community-based validation will be performed.

4.1. Gender and sexual diversity glossary

The Gender and Sexual Diversity Glossary is a bilingual (French-English) digital resource containing 155 concepts related to gender and sexual diversity. It was developed by the Translation

Bureau of the Government of Canada in collaboration with various organizations and experts, including linguists. The glossary was first published in 2017, in response to the inclusion of the concepts of “gender identity” and “gender expression” in the Canadian Human Rights Act and the Criminal Code. It was subsequently updated twice, in 2019 and in 2023, to reflect ongoing changes in trans-related vocabulary and community feedback [2]. This project draws on the most recent version, published in 2023 and available online.

4.2. Trends corpora

Trends corpora [52] are a family of large-scale monitor corpora available in Sketch Engine [53]. They are composed mainly of news articles and similar online resources, collected via RSS feeds, updated at regular times and grow rapidly [52, 54, 55, 56, 57, 58], although the time frames covered by individual languages differ (see Table 1). Each corpus is automatically lemmatized and part-of-speech tagged with language-specific tools and tagsets [52, 54, 55, 56, 57, 58]. Below is a table detailing the corpus info for each language in the project:

Table 1

Overview of Trends corpora [52] for the languages studied (January 2026). Data retrieved from Sketch Engine [53] and supplemented with corpus documentation [52, 54, 55, 56, 57, 58]

| | English | Polish | Spanish | French | Italian |
|--------------------|-----------------|---------------|---------------|---------------|----------------|
| earliest text date | 2014 | 2021q1 | 2021q1 | 2020q2 | 2014 |
| tokens | 102,290,203,390 | 1,322,725,117 | 4,068,431,653 | 1,703,314,501 | 12,008,227,395 |
| documents | 264,528,009 | 3,046,499 | 8,306,254 | 2,938,204 | 33,746,422 |
| regular updates | yes | yes | yes | yes | yes |

5. Discussion

5.1. Selection criteria and their limitations

Sourcing the initial set of units from an institutional, rather than a community-developed, resource introduces several challenges.

First, there might be a risk of “top-down bias”: categories and units may reflect official norms or ideological positions rather than community realities. However, one of the purposes of the tool is to confront this bias by empirically looking at the usage of the terms listed in institutional glossaries. Moreover, all efforts will be made to reduce this bias when defining the concepts by having all entries validated by community experts.

A related limitation is the fact that the final selection of units to be studied relied on the researcher’s intuition. Nevertheless, the number of units is by no means “set in stone”: it can be expanded if we manage to secure additional workforce.

Another challenge concerns the perceived innovativeness of the units. Since the tool is intended to serve a large usage base, the perception of novelty will be extremely variable, therefore we decided to also include terms which are not very recent, but which could be considered as such by users. This aspect is in line with the literature on social dynamic processes of neologisms [59, 60,

61, 62]. Additionally, some terms can be known formally, but their meaning is not widely known in the larger community.

The presumed centrality of the selected units for understanding trans experiences is open to debate. However, this aspect will be discussed at a later stage, when the communities are involved.

Finally, like many comparable resources, the glossary developed under this project treats each lexical unit as an independent entry and does not systematically account for its word family, i.e. all derived forms of a base word that are semantically related [63]. Although extending the glossary to include such forms would be highly beneficial, it raises a fundamental methodological issue: what principles would guide inclusion of a unit in a word family? This aspect becomes especially salient in the case of trans vocabulary characterized, as we have already pointed out, by rapid innovation and community-driven usage. For this reason, at this stage we focus on single entries; however, we consider expanding this aspect in the future.

As the literature review has shown, currently, there is no shared framework for identifying trans vocabulary. Community-based or institutionally-produced linguistic resources do not make their selection criteria transparent. As there is no similar cross-linguistic work, an explicit procedure had to be developed. Initially, each researcher identified interesting terms independently, but this approach proved inconsistent. As a result, we adopted the procedure described (Appendix B) to ensure coherence and comparability across languages.

5.2. Concept definition and domain classification

Although both the concept definition (what the unit denotes) and the domain classification (the area of social life in which the unit is used) were informed by existing resources (primarily the Gender and Sexual Diversity Glossary) and supported by corpus evidence, challenges remain. These include, among others, potential normative framing of categories, potential mismatches between institutional definitions and community-based usage, and potential bias in the corpus evidence itself. The corpus is made up of mainly news articles and is not trans community-oriented, which could result in a representation that does not fully capture community-based usage and meanings. However, the study of the press remains fundamental for capturing neologisms and emerging language varieties, as journalistic discourse plays a key role in the dissemination and vulgarisation of terms to a broader public.

To address these limitations and to strengthen the final concept definition, a second validation stage, involving consultations with community-based experts, is planned. Additionally, a study focusing on similarities and differences in concept definition across institutional and community-based resources is being designed. The results of this study will inform the final version of the concept definition implemented in the tool.

5.3. Corpora

A cross-linguistic corpus study of relatively new linguistic phenomena necessarily depends on the availability and suitability of the corpora.

Initially, we aimed for a large national reference corpora for each language (such as KWJP for Polish [64] or CORPES XXI for Spanish [65]). Nevertheless, preliminary searches proved that such corpora were not adequate for our purposes. The units under study were either extremely rare or absent. This is likely due to the fact that reference corpora tend to be dominated mostly by print media and trans vocabulary is mostly documented by digital, either community-based or institutional, resources and is used online, mostly in social media. Additionally, reference corpora are not usually updated on a regular basis, which does not make them suitable for studying rapidly evolving vocabulary related to trans experiences.

The Trends corpora [52] were chosen because of their Internet-based nature, regular updates, and relatively similar time frame of compilation across the languages studied. While these are clear advantages, this choice has its limitations. Because our focus is on European varieties of the

languages studied, we adopted a restrictive approach to the query protocol (Appendix B), excluding data from other regional varieties when necessary.

Furthermore, differences in the time frames covered in the Trends corpora across the languages studied must be taken into account when interpreting cross-linguistic frequency patterns.

Finally, although initially lemma-based searches were adopted due to the inflectional complexity of Polish, where grammatical case and other morphosyntactic features produce a wide range of word forms, they did not prove ideal for the lexical innovations under study. Built-in morphological analysers may fail to generate lemmas for items not included in their dictionaries, which leads to incomplete results. To address this limitation, we adopted a standardized wildcard procedure, which should compensate for this gap and produce more comprehensive results (Appendix B).

5.4. Three illustrative cases of terminological difficulties

The very nature of the lexis we intend to capture in this project poses specific challenges. Below, we outline three scenarios that illustrate typical difficulties we encountered when working on a terminological resource based on rapidly-evolving, community-driven and largely non-standardized vocabulary. Interestingly, these scenarios are common, although not to the same extent, for all languages studied. Each of them highlights a different type of limitation faced when working with corpus evidence.

The first scenario concerns lexical items that can be easily identified in the corpus in contexts related to trans experiences, but whose frequency is extremely low, limiting the possibilities for a more detailed description of their usage. An example is “agender”. In all the languages studied, its frequency was around 0.01 occurrences per million tokens. “Agender” mostly appears together with other non-binary gender identities (i.a. “pangender”, “bigender”), typically as part of a broader enumeration of such identities. This pattern, observed across the languages analyzed, may reflect the innovativeness of the unit and its limited diffusion in general language. At the same time, this extremely low number of attestations makes it challenging to produce a detailed description of its usage. Nevertheless, these findings also speak by themselves: low attestations and cooccurrence with other marginalized gender identities proves such terms occupy a peripheral place within this discourse and thus confirm the need to include them in the tool.

The second scenario represents the opposite situation. It concerns lexical items that might be treated as examples of semantic change [66], i.e. existing words or multi-word expressions that acquired a new, trans-related meaning. In such cases, the search typically yields an extremely large number of corpus hits not restricted to the trans-related domain. To address this challenge, we applied the filtering for relevance procedure (Appendix B, 3), which allowed us to limit the results to contexts relevant to our purposes. An example of such a scenario is “gender”. Across the languages analyzed, this unit produces a very high number of hits, and even after limiting them for relevance, these data remain too large to be comprehensively analyzed across all relevant contexts. In such cases, the final description of the unit is based on the collocational patterns and, if needed, on a random sample of occurrences.

Finally, the third scenario concerns units that are relatively easy to find in desired contexts, yet they are difficult to describe in terms of their discursive framings. Such a case would be “gender dysphoria”. The collocations of the expression, such as “rapid-onset”, “diagnosis” and “medically” mainly belong to the medical domain. This might be explained by the fact that, in many countries, a diagnosis of gender dysphoria is necessary for trans people to access healthcare. However, in 2019, the World Health Organization removed gender dysphoria from the list of mental and behavioural disorders. This medical- and therefore stigmatising-portrayal raises questions about the representation of terms and the trans experiences they denote in our corpus and, consequently, about the potential use of our tool.

6. Conclusions

The intended comprehensiveness and multilingual scope of the tool developed are both its greatest strengths and challenges. Throughout the project, methodological choices were guided by the tool's two-fold objective: being a resource useful for both linguists and the general public. Achieving this end required navigating conceptual, methodological and technical constraints, many of which are rooted in the nature of the object of the study: rapidly evolving and community-driven trans vocabulary.

In the course of the project, we identified several promising lines for future studies.

The rare presence of certain selected units in established reference corpora is not only telling of the nature of trans language, but also, in a broader sense, raises questions about the adequacy of such corpora for studying linguistic innovation which, nowadays, takes place mostly online, namely in social media. Comparative corpus studies examining the presence, frequency, and usage patterns of the units studied here in general-language corpora versus study corpora compiled with texts produced by trans people (e.g. social media posts, online forums) would certainly offer interesting insights.

These observations point toward the next steps in the development of this tool.

First, conducting the previously mentioned study comparing gender diversity-related concept definitions across institutional and community-based resources will inform the final version of the concept definition implemented in the tool and may reveal differences between normative vs non-normative framings of trans experiences.

Second, contrasting community-based usage of trans vocabulary with its usage by the general public and/or in specialized domains may shed light on how these concepts are framed across different types of discourse. Extending such studies across languages and cultures may further highlight culture-specific patterns.

Finally, developing a systematic procedure for the inclusion of word families in this type of digital terminological resource could prove applicable to studying and documenting the vocabulary of other non-normative communities.

Acknowledgements

This paper is based on work from COST Action ENEOLI, CA22126, supported by the European Cooperation in Science and Technology (COST), <http://www.eneoli.eu>.

Declaration on Generative AI

In the preparation of this paper, the authors used ChatGPT 5.2. for grammar and clarity check and language polishing. The authors carefully reviewed and edited the suggestions of the model and take full responsibility for the content of this paper.

References

- [1] European Network on Lexical Innovation (ENEOLI). 2023. URL: <https://eneoli.eu/>.
- [2] Government of Canada, Translation Bureau, Gender and Sexual Diversity Glossary, 2023. URL: <https://www.btb.termiumplus.gc.ca/publications/diversite-diversity-eng.html>.
- [3] Ministerio de Sanidad, Servicios Sociales e Igualdad, Glosario de términos sobre diversidad afectivo sexual. Documento de apoyo para el abordaje de la salud sexual y la diversidad afectivo sexual en la prevención de la infección por el VIH y otras ITS, 2018. URL: https://saludextremadura.ses.es/filescms/web/uploaded_files/CustomContentResources/GlosarioDiversidad.pdf.
- [4] Rada Języka Neutralnego, Słownik terminologii queerowej, n.d. URL: <https://zaimki.pl/terminologia>.

- [5] Istituto Superiore di Sanità (ISS), Ufficio Nazionale Antidiscriminazioni Razziali - Presidenza del Consiglio dei Ministri (UNAR), Glossario InfoTrans, 2020. URL: <https://www.infotrans.it/it-schede-8-glossario-transgender>.
- [6] Gouvernement du Canada, Bureau de la traduction, Lexique sur la diversité sexuelle et de genre, 2023. URL: <https://www.btb.termiumplus.gc.ca/publications/diversite-diversity-fra.html>.
- [7] A. S. Dyszak, Formy męsko- i żeńskoosobowe a niebinarni użytkownicy polszczyzny. U źródeł nowych zjawisk leksykalno-gramatycznych we współczesnym języku polskim, *Język Polski* 104 (2024) 32-47. doi:10.31286/JP.00965.
- [8] J. G. L. Lee, T. Ylioja, M. Lackey, Identifying Lesbian, gay, bisexual, and transgender search terminology: A systematic review of health systematic reviews, *PLoS One* 11 (2016). doi:10.1371/journal.pone.0156210.
- [9] R. Carabez, M. Pellegrini, A. Mankovitz, M. Eliason, M. Scott, Does your organization use gender inclusive forms? Nurses' confusion about trans* terminology, *Journal of Clinical Nursing* 24 (2015) 3306–3317. doi:10.1111/jocn.12942.
- [10] C. Míguez Bóveda, Perspectiva de género para todes: La necesidad de formación específica para traducir identidades no binarias a través del análisis de Gender Queer, *Asparkia. Investigació Feminista* 41 (2022) 115–137. doi:10.6035/asparkia.6070.
- [11] D. M. Tordoff, S. G. Haley, A. Shook, A. Kantor, J. M. Crouch, K. Ahrens, "Talk about Bodies": Recommendations for Using Transgender-Inclusive Language in Sex Education Curricula, *Sex Roles* 84 (2021) 152–165. doi:10.1007/s11199-020-01160-y.
- [12] S. Stryker, A. Aizura (Eds.), *The Transgender Studies Reader 2*, Routledge, New York, 2013.
- [13] T. Boellstorff, Playing Back the Nation: Waria, Indonesian Transvestites, *Cultural Anthropology* 19 (2004) 159-195. doi:10.1525/can.2004.19.2.159.
- [14] M. Rifkin, *When Did Indians Become Straight? Kinship, the History of Sexuality, and Native Sovereignty*, Oxford University Press, Oxford, 2011. doi:10.1093/acprof:oso/9780199755455.001.0001.
- [15] G. R. Bauer, R. Hammond, R. Travers, M. Kaay, K. M. Hohenadel, M. Boyce, "I Don't Think This Is Theoretical; This Is Our Lives": How erasure impacts health care for transgender people, *Journal of the Association of Nurses in AIDS Care* 20 (2009) 348-361. doi:10.1016/j.jana.2009.07.004.
- [16] D. Valentine, *Imagining Transgender: An Ethnography of a Category*, Duke University Press, Durham, 2007. doi:10.1215/9780822390213.
- [17] Y. G. Ansara, P. Hegarty, Cisgenderism in psychology: pathologising and misgendering children from 1999 to 2008, *Psychology & Sexuality* 3 (2012) 137-160. doi:10.1080/19419899.2011.576696.
- [18] J. Humbley, Socioterminology, in: J. Humbley, G. Budin, C. Laurén (Eds.), *Languages for Special Purposes: An International Handbook*, De Gruyter Mouton, Berlin, Boston, 2018, pp. 469-488. doi:10.1515/9783110228014-024.
- [19] L. Zimman, Transgender Language, Transgender Moment: Toward a Trans Linguistics, in: K. Hall, R. Barrett (Eds.), *The Oxford Handbook of Language and Sexuality*, Oxford University Press, 2018. doi:10.1093/oxfordhb/9780190212926.013.45.
- [20] P. Baker, R. Love, Corpus linguistics and sexuality, in: K. Hall, R. Barrett (Eds.), *The Oxford Handbook of Language and Sexuality*, Oxford University Press, 2018. doi:10.1093/oxfordhb/9780190212926.013.45.
- [21] K. Conrod, Pronouns raising and emerging (Linguistics), Ph.D. thesis, Univeristy of Washington, Washington, 2019. URL: <https://linguistics.washington.edu/research/graduate/pronouns-raising-and-emerging>.
- [22] S. Stryker, Transgender history, homonormativity, and disciplinarity, *Radical History Review* 100 (2008) 145-157. doi: 10.1215/01636545-2007-026.
- [23] L. Zimman, Transgender voices: Insights on identity, embodiment, and the gender of the voice, *Language and Linguistics Compass* 12 (2018) e12284. doi:10.1111/lnc3.12284.

- [24] M. T. Cabré Castellví, Une théorie multidimensionnelle des néologismes, *Neologica* 15 (2021) 25-41. URL: <https://classiques-garnier.com/neologica-2021-n-15-les-etudes-de-neologie-aux-xxie-siecle-un-etat-de-la-recherche-europeenne-une-theorie-multidimensionnelle-des-neologismes.html>.
- [25] N. Lo Vecchio, For a sociolexicology of queer naming: On problematizing the LGBTQ lexicon in historical linguistics, in: S. Burnett, F. Vigo (Eds.), *Battlefield Linguistics: Queer, Trans, and Feminist Interventions in Linguistic and Discursive Change*, De Gruyter Mouton, Berlin, 2026, pp. 157-178. doi:10.1515/9783111351568-008.
- [26] D. Knutson, J. M. Koch, C. Goldbach, Recommended terminology, pronouns, and documentation for work with transgender and non-binary populations, *Practice Innovations* 4 (2019) 214–224. doi:10.1037/pri0000098.
- [27] N. Thorne, A. K-T. Yip, W. P. Bouman, E. Marshall, J. Arcelus, The terminology of identities between, outside and beyond the gender binary – A systematic review, *International Journal of Transgenderism* 20 (2019) 138-154. doi:10.1080/15532739.2019.1640654.
- [28] A. Walkiewicz, Językowe wykładniki niebinarności płci w polszczyźnie. Część 1: Wprowadzenie do problematyki i próba systematyzacji, *Prace Językoznawcze* 24 (2022) 85–100. doi:10.31648/pj.7599.
- [29] M. Gębka-Wolak, Językowe wykładniki niebinarności płci w polszczyźnie. Część 2: Innowacje a system i tendencje rozwojowe, *Prace Językoznawcze*, 24 (2022) 101–116. doi:10.31648/pj.7602.
- [30] A. Franco Martínez, ¿Todos, todos/as, todxs o todes?: Efectos cognitivos del uso del genérico masculino y sus formas alternativas en español, in: M. Sánchez Ibáñez, M. Fernández Cano, A. Pérez Bernabeu, S. Fernández de Pablo (Eds.), *MariCorners: Investigaciones queer en la academia*, Egales, Madrid, 2019, pp. 25-56.
- [31] B. Papadopoulos, *Morphological Gender Innovations in Spanish of Genderqueer Speakers*, Bachelor's thesis, University of California, Berkeley, 2019.
- [32] M. Cabello Pino, Los morfemas de género emergentes (-x y -e) y su tratamiento en la prensa española, *Círculo de Lingüística Aplicada a la Comunicación* 89 (2022) 57-70. doi:10.5209/clac.79501.
- [33] C. M. de Onís, What's in an "x"?: An Exchange about the Politics of "Latinx", *Chiricú Journal: Latina/o Literatures, Arts, and Cultures* 1 (2017) 78-91. doi:10.2979/chiricu.1.2.07.
- [34] Á. López, Tú, yo, elle y el lenguaje no binario, *La Linterna del Traductor* 19 (2019) 142-150. URL: <https://www.lalinternadeltraductor.org/n19/traducir-lenguaje-no-binario.html>.
- [35] H. Lázaro Ruiz, Hacia la implementación del lenguaje inclusivo no binario en el aula de ELE: actitudes, usos, propuestas didácticas y retos, in: C. Soler Montes, M. Fuertes Gutiérrez (Eds.) *Gltopolítica de la diversidad. Una aproximación al lenguaje inclusivo en español. Educación Multidisciplinar Para La Igualdad De Género*, 6, Universidad Politécnica de Valencia, Valencia, 2025, pp. 99-124. URL: <https://monografias.editorial.upv.es/index.php/emig/article/view/614/545>.
- [36] C. Lizarbe Ruiz, *La terminología LGTBQ+ y de género: tratamiento, traducción y recursos disponibles*, Master's thesis, Universitat Oberta de Catalunya (UOC), Barcelona, 2018. URL: <https://openaccess.uoc.edu/server/api/core/bitstreams/7d6ea2fb-e086-4f31-83f6-514cb3adc41b/content>.
- [37] H. Cordova Quero, Hacia un breve glosario queer: algunas nociones acerca del género, la sexualidad y la teoría queer, *Análisis* 52 (2020) 95-121. doi:10.15332/21459169/5326.
- [38] R. L. Platero, M. Rosón, E. Ortega (Eds.), *Barbarismos queer y otras esdrújulas*, Edicions Bellaterra, Barcelona, 2017.
- [39] G. Guerrero Ramos, M. F. Pérez Lagos, Neologismos en el ámbito temático del colectivo LGTB, *Sabir International Bulletin of Applied Linguistics* 1 (2020) 143-176. doi:10.25115/ibal.v1i2.3712.
- [40] Moscas de Colores, Glosario LGBT. URL: <https://www.moscasdecolores.com/es/glosario-lgbt/>.
- [41] M. Solá García, *Guía básica sobre diversidad sexual y de género*, Gobierno de Navarra / Instituto Navarro para la Igualdad, 2020. URL: www.igualdadnavarra.es/imagenes/documentos/-235-f-es.pdf.

- [42] Alpheratz, Grammaire du français inclusif, Vent Solars, Châteauroux, 2018.
- [43] V. Swamy, L. Mackenzie (Eds.), Devenir non-binaire en français contemporain, Le Manuscrit, Paris, 2022.
- [44] K. A. Knisely, Le français non-binaire: Linguistic forms used by non-binary speakers of French, *Foreign Language Annals* 53 (2020) 850–876. doi:10.1111/flan.12500.
- [45] G. Sulis, V. Gheno, The debate on language and gender in Italy, from the visibility of women to inclusive language (1980s–2020s), *The Italianist* 42 (2022) 153–183. doi:10.1080/02614340.2022.2125707.
- [46] R. Baiocco, F. Rosati, J. Pistella, Italian proposal for non-binary and inclusive language: the schwa as a non-gender-specific ending, *Journal of Gay & Lesbian Mental Health* 27 (2023) 248–253. doi:10.1080/19359705.2023.2183537.
- [47] M. Rosola, M. Floris, D. Ruzzante, E. S. Safina, I. Facchini, G. Di Dona, G. Torrenzo, Double vowels, double fairness? Assessing the viability of diphthongs as novel strategies for gender fairness in Italian, *Language Science* 116, 2026. doi:10.1016/j.langsci.2026.101812.
- [48] Modena Pride, Glossario LGBTIQ. URL: <https://modenapride.it/scopri/glossario-lgbtqi>.
- [49] Torino Pride, Glossario LGBTQIA+. URL: <https://www.torinopride.it/glossario-lgbtqi/>.
- [50] Università di Trento, Le parole per includere: glossario LGBTQI+ UniTrento, 2024. URL: <https://www.unitn.it/sites/default/files/2024-11/Glossario%20LGBTQI%20%20UniTrento%202021.pdf>.
- [51] Università di Genova, Linee guida per l'utilizzo di un linguaggio inclusivo, 2023. URL: https://media.unige.it/media_production/Linee_guida_linguaggio_inclusivo_zstrv7.pdf.
- [52] O. Herman, M. Jakubiček, J. Kraus, V. Suchomel, From word of the year to word of the week: Daily-updated monitor corpora for 25 languages, in: *Electronic lexicography in the 21st century (eLex 2025): Intelligent lexicography. Proceedings of the eLex 2025 conference*, Lexical Computing CZ s.r.o., Bled, 2025, pp. 41–58. URL: https://elex.link/elex2025/wp-content/uploads/eLex2025-04-Herman_etal.pdf.
- [53] A. Kilgariff, V. Baisa, J. Bušta, M. Jakubiček, V. Kovář, J. Michelfeit, P. Rychlý, V. Suchomel, The Sketch Engine: Ten Years On, *Lexicography ASIALEX* 1 (2014) 7–36, doi:10.1007/s40607-014-0009-9.
- [54] Sketch Engine, English Trends corpus. n.d. URL: <https://www.sketchengine.eu/english-trends-corpus/#toggle-id-1>.
- [55] Sketch Engine, Polish Trends corpus. n.d. URL: <https://www.sketchengine.eu/polish-trends-corpus/>.
- [56] Sketch Engine, Spanish Trends corpus. n.d. URL: <https://www.sketchengine.eu/spanish-trends-corpus/>.
- [57] Sketch Engine, French Trends corpus. n.d. URL: <https://www.sketchengine.eu/french-trends-corpus/#toggle-id-1>.
- [58] Sketch Engine, Italian Trends corpus. n.d. URL: <https://www.sketchengine.eu/italian-trends-corpus/>.
- [59] R. Fischer, *Lexical Change in Present-Day English: A Corpus-Based Study of the Motivation, Institutionalization, and Productivity of Creative Neologisms*, Gunter Narr, Tübingen, 1998.
- [60] P. Hohenhaus, Lexicalization and institutionalization, in: P. Štekauer, R. Lieber (Eds.), *Handbook of Word-Formation. Studies in Natural Language and Linguistic Theory*, 64, Springer, Dordrecht, 2005, pp. 353–373.
- [61] D. Kerremans, *A Web of New Words. A Corpus-Based Study of the Conventionalization Process of English Neologisms*, Peter Lang, Berlin, 2015. doi:10.3726/978-3-653-04788-2.
- [62] H.-J. Schmid, *The dynamics of the linguistic system. Usage, conventionalization, and entrenchment*, Oxford University Press, Oxford, 2020. doi:10.1093/oso/9780198814771.001.0001.
- [63] L. Bauer, P. Nation, Word families, *International Journal of Lexicography* 6 (1993) 253–279. doi:10.1093/ijl/6.4.253.
- [64] W. Kieraś, M. Marciniak, M. Łaziński, M. Woliński, K. Bojałkowska, W. Eźlakowski, Ł. Kobyliński, D. Komosińska, K. Krasnowska-Kieraś, M. Rudolf, A. Tomaszewska, J. Wołoszyn,

- N. Zawadzka-Paluckta, Korpus Współczesnego Języka Polskiego. Dekada 2011–2020, *Język Polski* 105 (2025) 5-20. doi:10.31286/JP.001055.
- [65] Real Academia Española, Banco de datos (CORPES XXI). Corpus del Español del Siglo XXI (CORPES). n.d. URL: <http://www.rae.es>.
- [66] A. Blank, Why do new meanings occur? A cognitive typology of the motivations for lexical semantic change, in: A. Blank, P. Koch (Eds.), *Historical Semantics and Cognition*, De Gruyter Mouton, Berlin, Boston, 1999, pp. 61-90. doi:10.1515/9783110804195.61.

A. 30 units selected for inclusion in the tool

1. Agender
2. Bigender
3. Cisgender
4. Cisnormative
5. Cissexism
6. Cross-dressing
7. Coming out
8. Deadname
9. Detransition
10. Gender
11. Gender-affirming surgery
12. Gender assigned at birth
13. Gender binary
14. Gender-conforming
15. Gender diversity
16. Gender dysphoria
17. Gender expression
18. Gender identity
19. Gender neutrality
20. Gender nonconforming
21. Gender minority
22. Gender spectrum
23. Intersex
24. Misgender
25. Non-Binary
26. Pangender
27. Pronouns
28. Transgender
29. Transition
30. Transphobia

B. Query setup and note-taking

All researchers follow the same procedure.

1. Concordance search:

Concordance search is filtered by text type and top-level domains to include only domains that can be confidently identified as European in origin. This restrictive approach is aimed at controlling for regional variation in languages spoken in more than one country (English, Spanish, Italian, French).

All searches are conducted using basic concordance search in Sketch Engine. To address potential lemmatization problems, the asterisk (*) wildcard is used in all the queries. The asterisk (*) stands for zero or more characters. To retrieve all inflected forms of the Polish lemma *niebinarny* (non-binary), the query used is: *niebinarn**. This procedure allows for the retrieval of all relevant word forms, regardless of potential tagging inconsistencies.

2. Collocations:

If a unit generates more than 20 hits, the researcher examines collocations within a ± 5 -lemma window (KWIC), using lemma, not word form.

3. Filtering for relevance:

For units with broader or polysemous meanings, the analysis is restricted to concordances and collocations that are clearly related to trans-related issues. The procedure is as follows:

1. Open the *Collocations* tab.
2. Set the context window to ± 5 tokens (KWIC).
3. Use lemma-based collocations rather than word forms.
4. Select only those collocates that can plausibly be interpreted as related to trans matters.
5. Record the selected collocates in the *Comments* section of the shared working file.
6. Examine exclusively the concordances of the selected collocates.

4. Note-taking:

All observations are recorded in a shared working file. The researchers are asked to:

1. record the relative and normalized frequencies of the lexical unit at issue and all its variants;
2. collocations: used only for units with more than 20 occurrences; lists relevant lemma-based collocations;
3. comments/observations/usage: includes notes on usage patterns, verb constructions, semantic prosody, and other relevant distributional/contextual patterns;
4. additional remarks: information on the collocates selected for analysis in the case of broad-meaning units such as *gender*.