Liebe Kolleg:innen, querid@s compañer@s: presenting the GILDEES corpus

Marie-Pauline Krielke
Saarland University
Department of Language Science and Technology
Campus A2.2, 66123 Saarbrücken, Germany
mariepauline.krieke@uni-saarland.de

Abstract

We present a multi-register (WEB, NEWS, and GOVernment texts), diachronic (2015-2024), comparable corpus annotated for lexical gender-inclusive language (GIL) features in German and Spanish. Apart from rule-based annotations, we train a transformer-based classifier to resolve semantically ambiguous neutral expressions like epicenes to reliably annotate true human referents. In a sample study, we analyze register variation in the three registers in terms of GIL features both contrastively and diachronically. We show that GIL usage increases and varies diachronically in terms of register in both languages. German texts show a higher overall frequency and diversity of GIL features than Spanish texts. However, across languages, registers behave similarly, with government text showing the strongest usage of GIL followed by news and web texts, and web texts showing the strongest innovation in terms of features. The results of our study are valuable to linguistic areas such as human and machine translation, SLA, and contribute to register-conform gender inclusive NLP downstream tasks such as machine translation, summarization, or text generation. From a diachronic point of view, our corpus and analyses are a valuable contribution to observing language change in the making.

Keywords: Gender inclusive language, German-Spanish, contrastive corpus linguistics, diachronic change

1. Introduction

In recent years, gender-inclusive language (GIL) has become increasingly common in everyday usage, particularly in grammatically gendered languages such as German and Spanish. The equal representation of female and male referents has been debated in feminist linguistics since the 1970s in both German (e.g. Trömel-Plötz, 1978; Guentherodt et al., 1980; Pusch, 1979) and Spanish scholarship (Suardiaz, 1973; Hampares, 1976), and psycholinguistic studies on German masculine generics (e.g. Braun et al., 1998; Stahlberg et al., 2001; Gygax et al., 2008) show that they predominantly evoke male representations. However, the practical implementation of these findings in actual language use has been gradual and has achieved broader institutional uptake only in the past two decades, for instance through official guidelines (Günthner, 2019).

Since the early 2000s, the gender binary has increasingly been questioned, leading to proposals for including non-binary identities (Günthner, 2019), particularly through new orthographic forms (Diewald and Steinhauer, 2022) such as *Amigxs* or *Freund*innen*. As this development remains ongoing and socially contested, there is neither broad agreement on the use of GIL nor consensus on its lexical, grammatical, or orthographic realization.

It is therefore interesting to trace recent developments of GIL features, paying attention to the communicative contexts they are used in. While for German, a monolingual diachronic database of GIL in language use exists (Dick et al., 2024), it does not cover all lexical features of GIL, lacking, for instance, epicenes and collective terms, two strategies to refer to people without explicit gender marking. For Spanish, there is no comparable resource.

We present the GILDEES corpus, the first multi-register diachronic Spanish – German comparable corpus annotated for a comprehensive set of GIL features, i.e., explicit forms of gender inclusiveness such as spelling variants (*amig@*) and double mention (*amiga o amigo*), and implicit forms like epicenes (*persona*), collective terms (*equipo*), and nominalizations (*Studierende*).

A central part of this paper is dedicated to a detailed description of corpus building (Section 3) and annotation. Special focus is put on the annotation of implicit forms of GIL (Section 3.4), often ambiguous in their reference to personal or non-personal referents (e.g., *Haushaltshilfe*, domestic help, or the activity of helping in the household). To handle this kind of ambiguity, we manually annotate a gold standard set of epicenes and train transformer-based binary classifiers to disambiguate between personal and non-personal reference.

We present sample analyses (Section 5) to illustrate applications of the resource. Diachronically, Spanish lags behind German in both the frequency and variability of GIL features. Register patterns are similar across languages: government texts (GOV) show the highest GIL usage, web texts (WEB)

Proceedings of the 19th Workshop on Building and Using Comparable Corpora (BUCC) @ LREC 2026, pages 41–52
11 May 2026.

the lowest overall but the highest proportion of non-binary features, and news texts (NEWS) occupy an intermediate position with little feature variation. The GILDEES annotations thus enable tracking overall gender inclusiveness across registers as well as processes of diversification (increasing variability) and conventionalization (decreasing variability) over time, reflecting ongoing social negotiation and consensus formation around inclusive language.

2. Previous work

As awareness has increased that language can reflect and reinforce social inequalities and shape perceptions of gender and identity (e.g., Kaufmann and Bohner, 2014, for Spanish) and (Braun et al., 1998; Stahlberg et al., 2001; Gygax et al., 2008, 2019, for German), interest in research on the active usage of gender-inclusive language has grown substantially.

Practical implementations of GIL have been pursued through actively shaping language use via guidelines for a gender-inclusive usage in public discourse (e.g., Aguilar Gavira et al., 2019, for Spanish) and (e.g., Diewald and Steinhauer, 2022, for German). Especially in the area of gender-fair (machine) translation, research has been growing in the past few years (Daems, 2023; Piergentili et al., 2023; Lardelli et al., 2024; Savoldi et al., 2025).

Extant corpus-based research on GIL in German has shown that its usage underlies both temporal and register-dependent variation. In newspaper texts of the early 2000s, the generic masculine was still more prevalent than forms overtly including female referents (Bühlmann, 2002), while an integration of forms making both genders visible could be observed in job announcements (Demey, 2002). Well into the 2010s, an analysis of Swiss authorities texts shows a decrease in generic masculine compared to alternative forms such as double mentions (Elmiger et al., 2017). At the beginning of the 2020s, Sökefeld (2021) using a diachronic corpus (2000-2019) of newspaper and blog articles, reports the integration of overt non-binary GIL forms (*Freund*in*) from the early 2010s onwards, as well as a persisting preference for generic masculine forms in newspaper texts over alternative forms, while blog texts show a preference for neutral forms. Exemplifying the impact of political orientation in news texts, Rauth (2025) shows a temporal increase of GIL features in 2023 compared to 2021 in the German leftist newspaper *die Tagezeitung*. Political orientation also impacts GIL usage in the spoken public domain: Stecker et al. (2021) identify a general GIL increase in plenary protocols of the German Bundestag spanning 1949 - 2021 from the 1980s onward. Representatives of left-wing and green parties make stronger use of GIL than those of conservative parties. Regarding cross-register variation Dick et al. (2024) compile a multi-genre, diachronic corpus spanning 1993-2023 and find clear register-dependent differences (Twitter texts $>$ NEWS texts $>$ EU parliamentary text $>$ academic texts).

Research on GIL in the Spanish-speaking community began later than in the German-speaking community (Zapf, 2024). Similarly, there are fewer corpus-based studies tracing the temporal development of GIL usage. Qualitative studies (e.g Papadopoulos, 2022; Linares, 2022) reiterate the strong influence of the Spanish Royal Academy (RAE) hindering its adoption, especially regarding the recommendation to avoid overt non-binary GIL forms. Pino (2022), however, reports on a gradual increase in mentions of non-binary GIL features in press texts. In a pilot study using the Spanish reference corpus (CREA), Medina Guerra (2016) found an increase in neutral forms compared to generic masculine forms. Comprehensive resources and/or accounts on the diachronic development of GIL usage in Spanish, taking register variation into account, are still lacking.

To our knowledge, no comparable corpora currently exist for the German–Spanish language pair. For multilingual applications, the only related resource is a parallel dataset specifically designed to improve translations involving gender-neutral language (mGeNTE; Savoldi et al., 2025). However, even where corpus resources are available, the extraction and detection of GIL present substantial challenges. A major bottleneck concerns semantically ambiguous neutral forms, which may refer to both human and non-human entities, as well as grammatically ambiguous constructions such as nominalizations (Dick et al., 2024).

On the sociolinguistic level, the use of GIL is closely tied to speaker attitudes and ideological positioning (Greene and Rubin, 1991; Matheson and Kristiansen, 1987; Cremades and Fernández-Portero, 2022), and varies across demographic factors such as age (Parks and Roberton, 2008). Since such attitudinal and social factors are reflected in register-specific usage patterns, the creation of a multi-register resource is crucial for capturing the full variability of gender-inclusive language in practice.

3. Corpus Building

3.1. Corpus compilation

The corpus consists of a balanced number of texts covering the time span 2015-2024, which were crawled from the German and Spanish web, and covers three registers: blogs and forums (WEB),

Language	Register	Texts	Tokens
German	GOV	4,803	3,683,367
	NEWS	200,000	6,387,219
	WEB	5,270	5,990,239
Total		210,073	16,060,825
Spanish	GOV	5,150	5,436,287
	NEWS	200,000	8,591,051
	WEB	5,122	4,954,390
Total		210,272	18,981,728
Total		**420,345**	**35,042,553**

Table 1: Total corpus size by language and register.

governmental press releases (GOV), and news texts (NEWS). The WEB texts consist of blog articles from 10 different domains (psychology, cooking, gastronomy, travel, sports, tech, video games, lifestyle, finance, literature, education). For temporal attribution, they were selected according to their time stamps.

The NEWS texts consist of sentences taken from the downloadable Leipzig Corpora Collection (Goldhahn et al., 2012) and are derived from news feed texts. To control for regional differences, we restricted the dataset to sources from Spain and Germany, using DNS resolution of domain names to IP addresses. Geographic attribution of domains was obtained via the requests library (Reitz, 2016), which queried the public ipinfo.io API (IPinfo) to retrieve country-level metadata associated with resolved IP addresses. To make the sub-corpora's sizes roughly comparable, we took a random sample of 20,000 texts per language and year.

The government texts were semi-hand-crawled using BootCaT (Baroni et al., 2006). They were derived from nine Spanish and German governmental institutions, including the Ministries of Public Health, Ministries of Education, Offices for Migration, Ministries of Labor, the Police, Ministries for Family and Women, Ministries of Economy, Ministries of Finance, and the President's Offices.

The exact numbers of text and tokens per register are presented in Table 1.

3.2. Metadata

Each text in the corpus was annotated with metadata such as text ID, year as derived from timestamps, register (WEB, NEWS, GOV), URL, and author (i.e., newspaper, ministry, or blog title). WEB texts are additionally annotated with the topic of the blog, e.g., psychology, cooking, books, etc. Figure 1 shows the number of authors per register and per year, indicating that NEWS texts show the highest variability in authors with a decreasing trend. The WEB texts are relatively stable in the number of different authors per year and per language. GOV texts are completely stable with nine authors per language and year since their authors correspond to the different ministries.

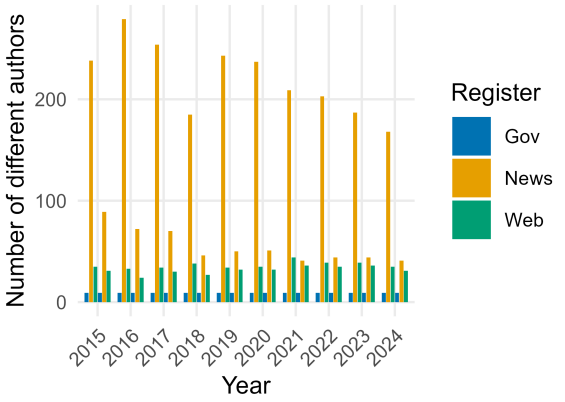

Figure 1: Authorial variation per register and year.

3.3. Morpho-syntactic annotation

Morpho-syntactic information was annotated using Stanza (Qi et al., 2020). The Spanish texts were parsed using the Stanza AnCora models to obtain language-specific XPOS tags. For German, we used Stanza with off-the-shelf settings. Morphological information is especially useful to detect gender-specific features. For instance, the double mention (amigos y amigas) can be detected via the morphological information.

3.4. Annotation of GIL features

The main focus of this work lies in the annotation of GIL features, which have been annotated using two different approaches: automatically with manual correction and manually with automatic generalization using machine learning approaches. Specifically, we divide GIL forms into explicit and implicit ones. Explicit GIL features are forms that explicitly mark the inclusion of other genders than merely the male. Implicit forms refer to those forms that do not encode natural gender and abstract away from any binary gender reference (Zapf, 2024, p.162). Explicit GIL is relatively easy to detect automatically, while implicit forms require manual approaches. These approaches will be described in the following. For an overview of GIL feature types, sub-types, and examples in Spanish and German, see Table 2.

3.4.1. Explicit GIL

Explicit GIL features are divided into two subcategories (Sökefeld, 2021): visibility and diversification. Visibility features refer to those visibly referring to both the male and the female gender (Zapf, 2024),

Type	Subtype	German	Spanish	Annotation
Explicit strategies				
Visibility	double mention	Bürger und Bürgerinnen	Los y las alumnos y alumnas	ex|vis|double
	infix-I	BürgerInnen	–	ex|vis|I
	grapho-stylistic disturbances	Bürger/innen, Bürger-innen, Bürger(innen)	alumnos/as, alumnos,-as, alumno(a)s	ex|vis|orth
	at-sign	–	alumn@s	ex|vis|at
Diversification	asterisk	Bürger*innen	alumn*s	ex|div|star
	underscore	Bürger_innen	alumn_s	ex|div|uscore
	colon	Bürger:innen	–	ex|div|colon
	-x	Bürgx	alumnxs	ex|div|x
	-e	–	alumnes	ex|div|e
Implicit strategies				
Neutralization	nominalization	Studierende, Beschäftigte, Arbeitslose	–	imp|neut|nom| (adj/part/num)
	derivation	-schaft, -hilfe, -kraft	–	imp|neut|der
	comunes	–	estudiantes, alumnado	imp|neut|com
Neutral/abstract forms	epicene	Person	persona	imp|epi
	collective terms	Team	equipo	imp|col

Table 2: Gender-inclusive linguistic strategies in German and Spanish (explicit vs. implicit).

most explicitly by mentioning both forms (double mention, e.g. *amigos y amigas*, coordinated by conjunctions like *o*/ *oder*, *y*/*und*, *bzw.* etc.), or using an infix-I (e.g., *FreundInnen*) or several forms of "grapho-stylistic disturbances" (cf. Gautherot, 2017, p.43) word-internal punctuation, e.g., the slash or the @-sign (*Schüler/innen*, *alumno/as*, *alumn@s*). Their explicit encoding facilitates automatic, rule-based annotation. We include nominal forms, but also pronominal forms (*Keine/r*, *niguno/a*) and determiners (*der/die*, *ein/-e*, *l@s*, *el/la*, etc.).

3.4.2. Implicit GIL

Implicit GIL forms are further divided into the subcategories neutralization and neutral/abstract forms (Bühlmann, 2002). **Neutralizations** refer to gender-neutral lexemes actively created (Gautherot, 2017; Sökefeld, 2021), such as derivations containing -kraft, -schaft, -hilfe, e.g. *Bürger-schaft* (citizens), *Haushalts-hilfe* (domestic help), *Führungs-kraft* (manager), and nominalizations derived from participles, e.g. Studierende (students), Geflüchtete (refugees), adjectives, e.g. Alte (elderly), or numerals, e.g. Hunderte (hundreds). These forms are highly productive in German; however, only inclusive in the plural form. In Spanish, this type of word formation is not inclusive since nominalizations are always gender-marked. Comparable forms in Spanish are the so-called comunes, i.e., lexemes lacking an explicit gender ending like *-a* or *-o*, e.g., *estudiantes* when intentionally used without a gender marked article (e.g. *Estudiantes de matemática se deben presentar a clase*) and were annotated following a manually curated list. Detection of nominalizations in German can be facilitated using UD-morphological annotation (de Marneffe et al., 2021). Plural nouns annotated without gender specification are possible candidates for nominalizations. These were extracted and manually cleaned for noise. Since not all lemmas encountered in this way were uniformly annotated without gender, in a second step, the cleaned set of lemmas was annotated with the word formation base form (adjective/participle/numerals) and used as a lookup list to automatically annotate all matching lemmas in the corpus.

Neutral/abstract forms include epicenes and collective terms (Bühlmann, 2002). Epicenes are inherently neutral expressions referring to persons, e.g., Person. However, many of them are semantically ambiguous, e.g. *Besuch* (visit or visitor). Collective terms in their most narrow definition refer to groups of people, such as *Team* or *equipo*. However, in the literature, abstract nouns referring to professions or positions (e.g., *consejo municipal*) or ministries (*Innenministerium*) are often also regarded as collective terms, referring metonymically to the group of people working within these institutions (Zapf, 2024). They represent the most problematic group, since most of them, apart from refer-

ring to groups of persons, can refer to institutions in their legal form, or buildings. Also, derivations with *-kraft* and *-hilfe* are semantically ambiguous, i.e. do not always refer to a person (e.g. *-hilfe* can refer to a person as in *Haushaltshilfe* but also to a technical device *Gehhilfe* (walking aid). The annotation guidelines are available on GitHub[1].

3.5. Personal reference resolution

To facilitate reliable identification and annotation of "true" derivations, epicenes, and collective nouns, linguistic experts compiled manually curated lemma lists. The German list contains both standalone lemmas and lemmas that may appear as compound constituents (e.g., -besuch), with compound status explicitly marked. Each lemma was further annotated for semantic ambiguity. The curated lists were subsequently used to automatically retrieve lemma occurrences from the corpus (including German compounds), and all unambiguous instances were automatically labeled accordingly (see Table 2). Ambiguous cases were annotated with the label "check" and held back for manual disambiguation in their context (see Section 3.5.1).

In a pilot trial, we used a machine learning approach to disambiguate automatically between epicenes referring to a person or not. For this, sentences with unique epicene lemmas previously identified as ambiguous, were extracted from the corpus. For each ambiguous lemma, we took a sample of at least one and up to 10 random sentences containing it, resulting in 2140 sentences with 716 unique lemmas for German. For Spanish, we allowed a sample size of 50 for each unique lemma since there are no compounds in Spanish. The sample resulted in 1145 sentences with 31 unique lemmas. Due to the great diversity of lemmas in German, the list of unique lemmas was divided into groups of ambiguity level (mostly person, 50/50 person/non-person, mostly non-person). To create the gold-standard training data, we down-sampled the 2140 sentences to 50% for manual correction, preserving the representativeness of the ambiguity groups. The Spanish set was left unchanged. Both sets were manually disambiguated and annotated with a binary label, e.g., imp|epi|person=yes or imp|epi|person=no, resulting in a total number of 1117 gold-annotated sentences with a distribution of 578 person=yes and 539 person=no for German and 1145 sentences for Spanish with 474 person=yes and 800 person=no showing a strong bias towards non-personal references.

3.5.1. Binary classifier

We implement a span-level transformer-based classifier that encodes full sentential context while restricting the classification decision to the contextualized representation of a manually annotated target expression, i.e., ambiguous epicenes. The model was implemented in Python 3 using PyTorch (Paszke et al., 2019). Pretrained transformer models and tokenizers were accessed via the Hugging Face Transformers library (Wolf et al., 2020), specifically the base version of XLM-RoBERTa (Conneau et al., 2020), since it is a multilingual model and can thus be used for both German and Spanish. The model was trained for six epochs with batch size 16, using cross-entropy loss and the AdamW optimizer (learning rate 2e^{-5}). Evaluation metrics (macro-averaged F1 and confusion matrix) were computed with scikit-learn (Pedregosa et al., 2011). Experiments were conducted in Google Colab with GPU acceleration via CUDA. The training data and code are available on GitHub[2].

Each instance consisted of the full sentence with explicit $< TARGET >$ epicene$< /TARGET >$ markers, enabling the model to attend explicitly to the referentially ambiguous token. Inputs were truncated or padded to a maximum sequence length of 192 tokens. Evaluation was conducted with predefined train (80%), development (10%), and test (10%) splits. For German, due to the high lemma diversity (many different compounds with huge overlap in head nouns), we constructed a strict lemma-held-out test set. On the held-out test set, the model achieved a macro-averaged F1 score of ≈ 0.90. The confusion matrix ($TN = 51$, $FP = 10$, $FN = 1$, $TP = 47$) indicates a slight asymmetry in errors, with more not_person instances misclassified as person (10 cases) than the reverse (1 case). At the same time, the model demonstrates particularly high recall for person-referential instances. Manual inspection of the false positives showed that, especially in cases where humans would also have struggled, led to wrong model predictions, e.g., compounds containing the lemma *Figur*, e.g., Spielfigur (play figure), which are ambiguous even in context. The Spanish model was trained with the same hyperparameters but without a lemma-held-out restriction, as the number of unique lemmas in the gold dataset was comparatively small. In Spanish, nouns are frequently modified by prepositional phrases rather than forming compounds (e.g., *personalidades del ámbito cultural*, i.e., personalities in the cultural sphere), which provide informative contextual cues for classification. The Spanish model performed slightly better than the German model, achieving a macro-F1 of ≈ 0.92 and demonstrating robust performance with minimal false negatives

[1] https://github.com/MariPeKa/GILDEES

[2] https://github.com/MariPeKa/GILDEES

for person-referential instances.

To assess practical usability, we simulated a selective prediction strategy in which thresholds of prediction probability were used to estimate the recall in automatic annotation. With restriction to $p \geq 0.8$ for $person = true$ and $p \leq 0.2$ for $person = false$, for Spanish, this approach achieved 93.7% automatic coverage, with a manual review rate of 6.3%. Within the automatically labeled subset, performance was highly reliable (precision ≈ 0.91), suggesting that model uncertainty effectively identifies borderline cases. Applying the same confidence-based filtering strategy to the German model resulted in 91.7% automatic coverage, with 8.3% of instances deferred for manual review. Among auto-decided cases, precision for *person* predictions was 0.849. Notably, recall for person-referential instances among automatically classified cases reached 1.00, indicating that no high-confidence person instances were missed.

Final automatic annotation in the corpus was therefore restricted to high-confidence predictions ($p \geq 0.8$ or $p \leq 0.2$), while ambiguous cases were labeled `check` and scheduled for manual review. Given the effectiveness of this approach for epicene annotation, we plan to extend it to collective terms in future work.

4. Access and Usage

The corpus is currently available in a derived format[3] to circumvent copyright restrictions on the complete web and government texts. In the latter case, we removed all sentences without GIL features and retained only those containing GIL features. Of the remaining sentences, we masked the word forms and lemmas of all but the GIL forms and their 4 preceding and 5 following words (see example 7). The full corpus may only be made available to specific individuals for the purpose of reviewing my research work.

5. Sample analysis

To examine the temporal development of GIL features in both languages, we conduct quantitative analyses based on all previously annotated, unambiguous features according to the following hypotheses. We first analyze overall feature frequencies. We then compare feature variability across languages and compute yearly feature diversity using entropy.

Hypotheses

1. Contrastive dimension (H1): German texts exhibit earlier and stronger adoption of GIL features.

2. Register dimension (H2): GIL usage frequencies are highest in GOV, followed by NEWS and WEB.

3. Diachronic dimension (H3): GIL usage increases across languages and registers.

4. Variability (H4): (a) German shows greater feature variability than Spanish. (b) WEB texts show the highest feature variability among all registers.

5.1. GIL over time per register in German and Spanish

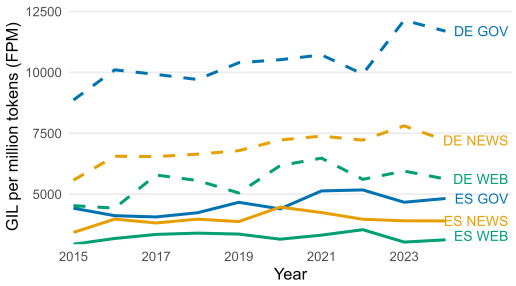

Figure 2: GIL frequencies in ES and DE per year and register.

Normalized frequencies of GIL features per language year and register, reveal both language- and register-specific differences (Figure 2).

For German, we observe consistently higher frequencies than in Spanish (H1). GOV and NEWS texts show a clear upward trend, both peaking in 2023. As expected, WEB texts show the lowest usage of GIL features overall, with stronger oscillations and an earlier peak in 2021.

In the Spanish corpus, the order of frequencies is the same, with GOV texts showing the strongest GIL usage, and an increasing trend. They are followed by NEWS and WEB texts.

The similar proportions per register are plausible: Government press releases rely most heavily on personal references when addressing topics of public interest. NEWS texts show an intermediate picture reporting about recent events and mostly specific persons of public interest whose gender identity is mostly known, and abstraction/inclusion is not needed. WEB texts show the lowest amount of (GIL) as blogs and forums often represent reports written from a first-person perspective, their experiences and opinions, rather than referring to others. The results are in line with H2.

[3] https://zenodo.org/records/19236744

Diachronically, we only see a consistent increase in GIL in German, while in Spanish, GIL usage only increases in GOV texts, partially confirming H3.

5.2. GIL feature diversity

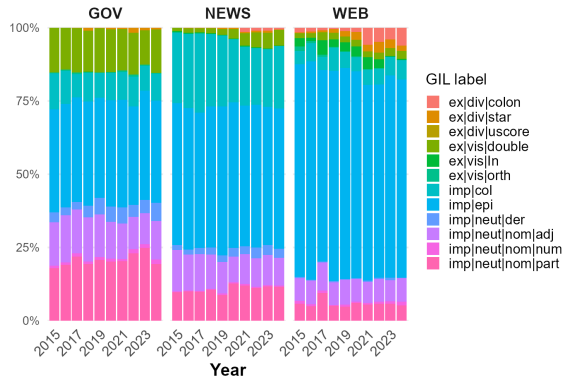

Figure 3: GIL types in German across years and registers.

German The distributions of GIL features differ across registers (Figure 3): Overall, implicit GIL features show the highest proportional representation in all registers, with epicenes being the most represented feature. This is not surprising since epicenes (e.g. *persona*) are highly frequent lexemes of personal reference in any register. Their dominance is especially high in the WEB register. A look into the WEB subcorpus reveals that especially the lemma *Kind* is highly frequent, describing a frequent topic of interest in the blogs. In NEWS texts, collective terms take the second-highest proportion, plausibly, since NEWS texts frequently report on governmental activities and their impact on the citizens. Frequent lemmas are *Regierung* (government) and *Bevölkerung* (population). Neutralizations, especially nominalizations derived from participles and adjectives, are the second most prevailing group in GOV and WEB texts. Explicit features take a lower proportion in all registers.

GOV texts show a high proportion of double mentions (e.g., *Bürgerinnen und Bürger*), a long-established and comparatively uncontested GIL strategy in German that explicitly addresses women and men alike. Their distribution remains stable over time, suggesting institutionalized use in public discourse. The only other explicit feature is the gender star, attested from 2018 onward (*Migrant*nnen*). GOV exhibits the most even and diverse distribution of GIL features across registers, with minimal temporal variation, indicating register-specific consolidation.

NEWS texts exhibit a relatively stable distribution dominated by implicit features. Among explicit forms, only double mentions increase in frequency, along with a small share of colon forms (*Bürger:innen*), which peak in 2021.

WEB texts show the highest and constantly increasing proportion of features of diversification, most prominently the colon (*Musiker:innen*) and the gender star (*Trainer*innen*). Of all registers, WEB is the only register with a notable proportion of using the infix-I (*AutorInnen*). The corpus data shows that the infix-I is also often used in hybrid form, e.g., in combination with a slash (*Follower/Innen*). Compared to the other registers, the proportion of double mentions is comparatively low. The diversity of the explicit GIL features increases over time and is biggest compared to the other registers. This is in line with our assumptions (H4b), as authors of WEB texts have the biggest freedom to linguistic innovation and diversity due to lower editorial regulation.

From a probabilistic point of view, the three registers differ in terms of overall entropy (Figure 4), which is highest for GOV texts, in line with the most even distribution of features in this register. The high entropy also indicates that German institutional texts make a balanced use of available GIL features. The fairly stable trend in entropy points to conventionalization of an already established set of GIL choices with little temporal variation or innovation. For NEWS and WEB texts, entropy across features both increases due to a diversification of feature usage over time. This trend is in line with our expectations since the increasing social demand for gender inclusion in language seems to have pushed towards a stronger integration of explicit forms apart from already existing neutral features. The lowest overall entropy in WEB texts can be attributed to the strong dominance of epicenes in the register. It disguises, however, the great variability in explicit features.

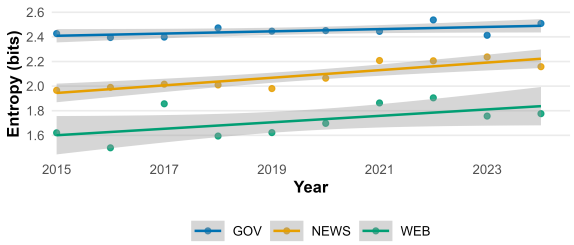

Figure 4: Entropy in German across years and registers.

Spanish For Spanish (Figure 5), we observe a more similar distribution of features across registers than in German, with epicenes being by far the most frequent ones.

GOV texts show the most visible changes in their GIL feature proportions over time. Starting out with

a high dominance of epicenes, their proportion decreases over time while both comunes (e.g. *habitantes*) and double mentions increase proportionally. Overall, the usage is restricted to rather conservative forms of GIL: neutral and visibility forms are represented, forms including gender diversity (beyond binary gender) are not used in official GOV texts in Spain.

NEWS texts show the lowest and a decreasing trend of variability in the set of GIL options used. The features used are very stable in temporal distribution with only a slight shift towards stronger epicene proportions. This stability is especially remarkable considering the high authorial variability in the NEWS register (cf. Figure 1).

WEB texts are fairly stable in their proportions of epicenes and comunes. As in the German WEB texts, epicenes show the highest proportions compared to the other registers (+50%). Also, similar to the German texts, WEB shows low but comparatively the largest proportions of explicit GIL features across registers, but with strong oscillations, reflecting more heterogeneous authorship, with personal choices of GIL rather than a concerted strategy of gender-inclusiveness. Counter to our intuition, the diversification feature x (*amigxs*) is practically not used. Also, the use of @ (*amig@* seems to decrease over time. The visibility features double mention and orthographical features like the slash (*amigos/as*) show opposite trends: double mentions proportion increases while orthographical features decrease.

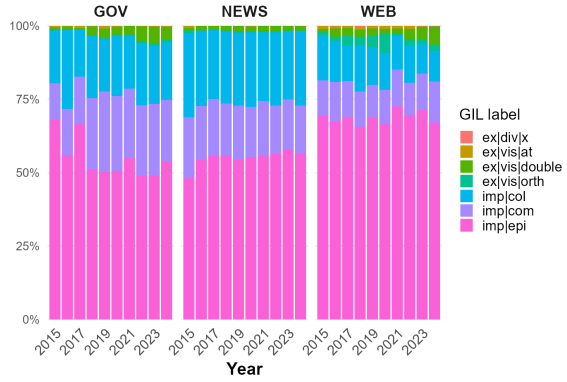

Figure 5: GIL types in Spanish across years and registers.

In probabilistic terms, the Spanish GOV texts show a notable increase in entropy, while in NEWS texts, entropy decreases (Figure 6). The increase in the GOV texts reflects the decreasing bias for epicenes, indicating an ongoing diversification with different ways of using GIL in the public discourse, possibly influenced by the change in administrations from the conservative People's Party (PP) to the left-wing Spanish Socialist Workers' Party (PSOE) in

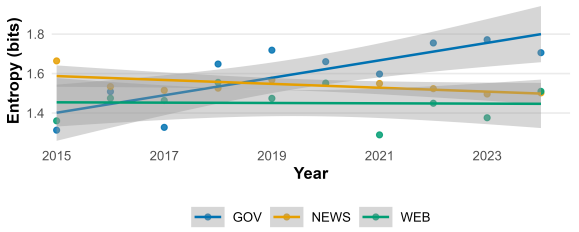

Figure 6: Entropy in Spanish across years and registers.

2018. The decrease in variability in NEWS texts is driven by a shift towards a stronger preference for epicenes and indicates conventionalization. The strong scattering of entropy values across the years in the WEB texts is an indicator of an ongoing experimentation with GIL options and temporal oscillations in preference for a specific option.

Comparing the developments in German and Spanish, we can conclude that the GOV and WEB texts in both languages behave similarly. In both German and Spanish, GIL usage is highest overall and shows the highest variability. The German GOV texts, however, start at higher frequencies and higher entropy rates, while the Spanish texts show a much milder increase in both over time. This reflects our assumptions about the time-shifted development of Spanish GIL usage compared to German. Both German and Spanish WEB texts show the highest usage of epicenes and explicit GIL features, including diversification strategies representing non-binary gender identities.

6. Conclusion

We have presented a diachronic, comparable German-Spanish corpus annotated with explicit and implicit GIL features. We have described corpus compilation and annotation, especially that of GIL features. We discussed the problem of ambiguity with collective and epicene terms and proposed a binary classification technique using a transformer model to solve ambiguous cases, achieving a classifier accuracy of +90%. The successful application of this method serves as a motivation to apply it to collective terms in future work.

A sample analysis was conducted to trace GIL usage cross-linguistically and diachronically per register. Our assumption that German spearheads GIL usage, displaying a higher GIL usage compared to Spanish, was confirmed. We also confirmed the assumption that over the past ten years, there is a diachronic GIL increase overall in both languages. In terms of register, we showed that cross-linguistically, GOV texts make the strongest use of GIL features. WEB texts display the highest number of different GIL features and highest proportions of

explicit GIL features involving grapho-stylistic disturbances such as *amig@* or *Freund*innen* facilitating explicit reference to (non-) binary gender identities.

The GILDEES resource is a valuable contribution to the study of GIL use in Spanish and German, but can also be used for other register-based and /or diachronic contrastive studies.

7. Acknowledgements

This research was funded by the Internationalization Fund of Saarland University. We thank the reviewers for their valuable comments and suggestions. We also gratefully acknowledge the annotators of the corpus: Maria Basova, Léon Maurice Jost, and Valentina Fajardo.

Limitations

Corpus annotation bears many difficulties. Apart from errors in automatic annotations, which we have tried to maintain at a minimum through effortful manual annotation and correction work, attributions of ambiguous cases are the biggest source for errors. There are cases in which even a human annotator would struggle to decide whether the term refers to a person (Example 4), or whether the label epicene (referring to a single person) or collective (referring to a group of people) would be more applicable. While Example 1 is unambiguously refers to an activity, and Example 2 refers unambiguously to a person, Example 3 stays ambiguous regarding the type of personal reference (individual or collective).

1. *Ich habe heute **Aufsicht** in der Pause.*
   *EN: I am doing **recess supervison** today.*
   (non-person).

2. *Die **Aufsicht** geht den Gang entlang.*
   *EN: The **supervisor** walks down the hallway.*
   (single person).

3. *Die **Aufsicht** signalisierte Zustimmung.*
   *EN: The **supervisory authority** signaled its approval.*
   (ambiguous: collective person/single person).

4. *Wir haben das Rundum-Sorglos-Paket: Mittagessen, **Hausaufgabenbetreuung**, sauberes Haus.*
   *EN: We have the all-inclusive-carefree package, lunch, **homework support**, clean house.*
   (person?)

Another limitation lies in deciding whether general reference to persons using epicenes and collective terms is intentional GIL usage. The assumption that GIL is used to linguistically represent more than just male gender identities implies a certain degree of intentionality. With inherently neutral forms like epicenes and collective nouns, this intentionality is impossible to prove and needs to be inferred in context. For the present study, we follow the axiom that all neutral references represent inclusive ways of reference, irrespective of the author's intention, and therefore count them as GIL features. Their unknown intentionality status, however, suggests that their inclusiveness operates on a different scale. We intend to address this issue in future, more conceptual work.

Finally, due to copyright restrictions, the corpus in its original textual form cannot be made publicly available and may only be shared with designated reviewers for research evaluation purposes; therefore, for legally compliant public dissemination, it is available in a derived format that preserves metadata and linguistic annotations while masking words which do not belong to the immediate context (preceding 4 and following 5 words) of the GIL forms prevent reconstruction of the original copyrighted texts.

8. Bibliographical References

Sonia Aguilar Gavira, Remedios Benítez Gavira, et al. 2019. Guía para un uso inclusivo del lenguaje.

Marco Baroni, Adam Kilgarriff, Jan Pomikálek, and Pavel Rychlỳ. 2006. Webbootcat. instant domain-specific corpora to support human translators. In *Proceedings of the 11th annual conference of the european association for machine translation*.

Friederike Braun, Anja Gottburgsen, Sabine Sczesny, and Dagmar Stahlberg. 1998. Können Geophysiker Frauen sein? Generische Personenbezeichnungen im Deutschen. *Zeitschrift für germanistische Linguistik*, 26(3):265–283.

Regula Bühlmann. 2002. Ehefrau Vreni haucht ihm ins Ohr... Untersuchung zur geschlechtergerechten Sprache und zur Darstellung von Frauen in deutschschweizer Tageszeitungen. *Linguistik Online*, 11(2).

Alexis Conneau, Kartikay Khandelwal, and Naman Goyal. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Raúl Cremades and Ignacio Fernández-Portero. 2022. Actitudes del alumnado universitario ante el lenguaje inclusivo y su debate en los medios de comunicación. 89:89–116.

Joke Daems. 2023. Gender-inclusive translation for a gender-inclusive sport: strategies and translator perceptions at the international quadball association. In *Proceedings of the First Workshop on Gender-Inclusive Translation Technologies*, pages 37–47, Tampere, Finland. European Association for Machine Translation.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Eline Demey. 2002. Leser und leserinnen gesucht! zum generischen gebrauch von personenbezeichnungen in deutschen stellenanzeigen und zeitungsartikeln. *Deutsche Sprache*, 30(1):28–49.

Anna-Katharina Dick, Matthias Drews, Valentin Pickard, and Victoria Pierz. 2024. GIL-GALaD: Gender inclusive language - German auto-assembled large database. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7740–7745. ELRA and ICCL.

Gabriele Diewald and Anja Steinhauer. 2022. *Handbuch geschlechtergerechte Sprache: Wie Sie angemessen und verständlich gendern*. Duden.

Daniel Elmiger, Eva Schaeffer-Lacroix, and Verena Tunger. 2017. Geschlechtergerechte sprache in schweizer behördentexten: Möglichkeiten und Grenzen einer mehrsprachigen Umsetzung. *Osnabrücker Beiträge zur Sprachtheorie*, 90:61–90.

Laure Gautherot. 2017. Vom Sprachfeminismus zum gendergerechten Sprachgebrauch in der BRD. *Kwartalnik Neofilologiczny*, (1):39–53.

Dirk Goldhahn, Thomas Eckart, Uwe Quasthoff, et al. 2012. Building large monolingual dictionaries at the leipzig corpora collection: From 100 to 200 languages. In *LREC*, volume 29, pages 31–43.

Kathryn Greene and Donald L. Rubin. 1991. Effects of gender inclusive/exclusive language in religious discourse. 10(2):81–98.

Ingrid Guentherodt, Marlies Hellinger, Luise F. Pusch, and Senta Trömel-Plötz. 1980. Richtlinien zur vermeidung sexistischen sprachgebrauchs.(principes pour éviter les usages sexistes dans la langue). *Linguistische Berichte Braunschweig*, (69):15–21.

Susanne Günthner. 2019. Sprachwissenschaft und Geschlechterforschung: Übermittelt unsere sprache ein androzentrisches weltbild? In *Handbuch Interdisziplinäre Geschlechterforschung*, pages 571–579. Springer.

Pascal Gygax, Ute Gabriel, Oriane Sarrasin, Jane Oakhill, and Alan Garnham. 2008. Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and cognitive processes*, 23(3):464–485.

Pascal Mark Gygax, Daniel Elmiger, Sandrine Zufferey, Alan Garnham, Sabine Sczesny, Lisa Von Stockhausen, Friederike Braun, and Jane Oakhill. 2019. A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Frontiers in psychology*, 10:1604.

Katherine J Hampares. 1976. Sexism in Spanish lexicography? *Hispania*, 59(1):100–109.

IPinfo. Ipinfo ip geolocation api. Accessed 2025-04-20.

Christiane Kaufmann and Gerd Bohner. 2014. Masculine generics and gender-aware alternatives in Spanish. *IZGOnZeit. Onlinezeitschrift des Interdisziplinären Zentrums für Geschlechterforschung (IZG)*, pages 8–17.

Manuel Lardelli, Timm Dill, Giuseppe Attanasio, and Anne Lauscher. 2024. Sparks of fairness: Preliminary evidence of commercial machine translation as English-to-German gender-fair dictionaries. In *Proceedings of the 2nd International Workshop on Gender-Inclusive Translation Technologies*, pages 12–21, Sheffield, United Kingdom. European Association for Machine Translation (EAMT).

M. ª Antonia Martínez Linares. 2022. Sobre los dobletes de género y cuestiones gramaticales conexas. 89:71–88.

Kimberly Matheson and Connie M. Kristiansen. 1987. The effect of sexist attitudes and social structure on the use of sex-biased pronouns. *The Journal of Social Psychology*, 127(4):395–398.

Antonia María Medina Guerra. 2016. Las alternativas al masculino genérico y su uso en el español de españa. *Estudios de lingüística aplicada*, 34(64).

Ben Papadopoulos. 2022. A brief history of gender-inclusive spanish. 48(1):31–48.

Janet B Parks and Mary Ann Roberton. 2008. Generation gaps in attitudes toward sexist/nonsexist language. *Journal of Language and Social Psychology*, 27(3):276–283.

Adam Paszke, Sam Gross, Francisco Massa, and et al. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32.

Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, and et al. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830.

Andrea Piergentili, Dennis Fucci, Beatrice Savoldi, Luisa Bentivogli, and Matteo Negri. 2023. Gender neutralization for an inclusive machine translation: from theoretical foundations to open challenges.

Manuel Cabello Pino. 2022. Los morfemas de género emergentes (-x y -e) y su tratamiento en la prensa española. 89:57–70.

Luise Pusch. 1979. Der Mensch ist ein Gewohnheitstier, doch weiter kommt man ohne ihn. Eine Antwort auf Kalverkämpers Kritik an Tromel-Plotz' Artikel über Linguistik und Frauensprache. *Linguistische Berichte Braunschweig*, (63):84–102.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108. Association for Computational Linguistics.

Philip Rauth. 2025. Freund:innenliche Chef:innen. Eine Korpusstudie zu Gendermovierungen. In Alexander Werth, editor, *Die Movierung: Formen, Funktionen, Bewertungen*, pages 332–364. De Gruyter, Berlin/Boston.

Kenneth Reitz. 2016. *Requests documentation*. EEUU: Requests Documentation.

Beatrice Savoldi, Eleonora Cupin, Manjinder Thind, Anne Lauscher, Andrea Piergentili, Matteo Negri, and Luisa Bentivogli. 2025. mGeNTE: A multilingual resource for gender-neutral language and translation. ADS Bibcode: 2025arXiv250109409S.

Dagmar Stahlberg, Sabine Sczesny, and Friederike Braun. 2001. Name Your Favorite Musician: Effects of Masculine Generics and of their Alternatives in German. 20(4):464–469.

Christian Stecker, Jochen Müller, Andreas Blätte, and Christoph Leonhardt. 2021. The evolution of gender-inclusive language. Evidence from the German Bundestag, 1949-2021.

Delia Esther. Suardiaz. 1973. *Sexism in the Spanish language*. University of Washington studies in linguistics and language learning ; v. 11. [Distributed by the Department of Linguistics, University of Washington], Seattle.

Carla Sökefeld. 2021. Gender (un) gerechte Personenbezeichnungen: derzeitiger Sprachgebrauch, Einflussfaktoren auf die Sprachwahl und diachrone Entwicklung. 46(1):111–141.

Senta Trömel-Plötz. 1978. Linguistik und Frauensprache.(linguistique et langage féminin). *Linguistische Berichte Braunschweig*, (57):49–68.

Thomas Wolf, Lysandre Debut, Victor Sanh, and et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*.

Miriam Zapf. 2024. *Gender, Sprache, Kognition: Eine linguistische Untersuchung zu genderinklusivem Sprachgebrauch im Spanischen*. Linguistik in Empirie und Theorie/Empirical and Theoretical Linguistics. Springer.

9. *Appendix A.*

```
<s id="23">
1→MASK→MASK→ADV→ADV→PronType=Ind→2→advmod→_→_
2→MASK→MASK→DET→ADV→Degree=Cmp|PronType=Ind→9→nsubj→_→_
3→als→als→ADP→KOKOM→_→5→case→_→_
4→die→der→DET→ART→Case=Nom|Definite=Def|Gender=Fem|Number=Sing|PronType=Art→5→det→_→_
5→Hälfte→Hälfte→NOUN→NN→Case=Nom|Gender=Fem|Number=Sing→2→nmod→_→_
6→der→der→DET→ART→Case=Gen|Definite=Def|Number=Plur|PronType=Art→7→det→_→_
7→Teilnehmenden→Teilnehmend→NOUN→NN→Case=Gen|Number=Plur→5→nmod→_→imp|neut|nom|part
8→sind→sein→AUX→VAFIN→Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin→9→cop→_→_
9→Männer→Mann→NOUN→NN→Case=Nom|Gender=Masc|Number=Plur→0→root→_→_
10→.→.→PUNCT→$.→_→9→punct→_→_
</s>
```

Figure 7: Derived corpus publication format