Andrea Abel and Natascia Ralli

GENDER IN ELECTRONIC DICTIONARIES AND TERMINOLOGY DATABASES State of the Art and Future Directions

Abstract What strategies are currently being applied in electronic dictionaries and terminology databases to gender representation, with a particular focus on feminine agentives? Starting with an overview of the state of the art as to gender studies in lexicography and terminology science, in this paper we use a case study to reflect upon the approaches in electronic dictionaries and terminology databases, collecting both good practices and concrete limitations that hinder an equal gender representation by contrasting different linguistic and socio-cultural contexts. Since lexical resources can be considered not only utility instruments but also "agents" that reflect society (Gouws, 2022, p. 40) there is a risk that they contribute, for example, to perpetuate gender stereotypes (Müller-Spitzer, 2023, p. 80). Thus, it is important for lexicographers and terminologists to discuss and establish methods for avoiding gender bias in lexical resources.

Keywords lexicography; terminology work; gender

1. Introduction

With the rise of the feminist movement in the 1960s, the attention towards all forms of sexism, including linguistic sexism, became an object of scientific and popular discussions. 60 years later, the debate has not dried up yet. On the contrary, in recent years, it has regained the general public's interest, especially through the #MeToo movement (Cavagnoli & Dragotto, 2021, p. 1) and an increasing devotion of the media to gender inequalities in general. Also, regarding language, a steadily growing number of questions have been posed by the public to academies' experts (e.g., Accademia della Crusca; RAE@informa).

In the field of lexicography, since the 1970s, numerous authors analyzed and reflected upon gender treatment in dictionaries (e.g., Graham, 1975; Hampares, 1976; Pusch, 1984; Whitcut, 1984; Calvo Ramos, 1998; Nübling, 2009; Westveer et al., 2018; Fuertes-Olivera & Tarp, 2022; Müller-Spitzer & Lobin, 2022). Discussions in this field have led some dictionaries to revise their approach, deciding to make the feminine agentives, i.e., "linguistic forms that indicate an agent, as job titles, etc." (Bengoechea. 2017, p. 200),¹ visible as separate entry or to record the masculine and feminine agentives on the same level (e.g., *Duden, Diccionario de la lengua española of the Real Academia Española, Oxford English Dictionary, Treccani*).

¹ The concept "agentive" also includes professions, jobs, roles, and educational qualifications.

In terminology work, the debate on how feminine agentives should be recorded is still open (Ralli & Evers, in review). The problem is methodological at base: concept is the core of terminology science and, as such, is not related to gender. However, there is the need to record feminine agentives in terminology databases to be able to use them correctly and consistently also in combination with, for example, proofreading tools, machine translation engines or computer-assisted translation (CAT) tools (Evers, 2022, p. 13).

Against this background, we want to address the following question in this paper: What strategies are currently being applied in electronic dictionaries and terminology databases to gender representation, with a particular focus on feminine agentives? We address this question by describing concrete examples from mainly two different socio-cultural contexts, the German-speaking area and the Italian-speaking area. We concentrate on the German dictionary Duden online and the Italian dictionary Treccani (online version) for the exemplary analyses of electronic dictionaries. The terminology databases, on the other hand, are LinguaPC, TERMDAT, and TERMCAT. Through this analysis, we pursue two main objectives. Firstly, we aim to summarize the state of the art. Secondly, we describe, by means of exemplary illustrations, the current strategies applied in electronic dictionaries and terminology databases to represent feminine agentives.² We concentrate on both their macrostructure (lemmatization) and their microstructure (definitions and examples), taking into account the different methodological approaches, a predominantly semasiological one in lexicography versus a predominantly onomasiological one in terminology work which leads to different foci.

2. State of the Art in Lexicography

2.1 German-Speaking Area

The well-known article "Sie sah zu ihm auf wie zu einem Gott – Das DUDEN-Bedeutungswörterbuch als Trivialroman"³ by Pusch (1984) is usually cited as the starting point for the discussion of gender issues in German lexicography. In it, Pusch examined part of the first edition of the dictionary from 1970 and found, for example, that only a small proportion of the personal names refer to women and that the role of women is strongly orientated towards men and children. Women are often portrayed either as virtuous and naive or as seductive and cunning (see Kotthoff & Nübling, 2018, pp. 180–181). In subsequent editions, the Duden editors responded – to varying degrees – to the criticism. The Duden editors themselves described in an article (Kunkel-Razum, 2012, pp. 2015–2016), in the title of which they referred directly to Pusch with "Er sah zu ihr auf wie zu einer Göttin",⁴ how the theoretical discussion in the context of gender studies was incorporated into practical dictionary work. In 2002, for example, female personal names were included in all entries for

² We have limited the analysis to binary forms.

 $^{^{}_{\rm 3}}$ "She looked up to him like a god – The DUDEN dictionary of meanings as a light novel" (translation by the authors).

⁴ "He looked up to her like to a goddess" (translation by the authors).

male personal names – with *Päpstin⁵* as the only exception (however, included in 2009⁶) – to make women more visible in dictionaries. More revisions followed in response to criticism by Porsch (2004), for example, in an article also alluding to Pusch, as "Das DUDEN-Universalwörterbuch als Fortsetzung eines Trivialromans"⁷ reveals in the title.

A comparative overview since the 1970s by Kotthoff & Nübling (2018, pp. 180–184) shows that stereotypical role models can also be found in German dictionaries from other publishers, with only the *Wörterbuch der deutschen Gegenwartssprache* (1961–1977) from the former GDR partially breaking through the usual stereotypes. More recent online dictionaries also reproduce traditional role models, as Thüne & Leonardi (2006) show using the *German Word Thesaurus*.

The fact that dictionaries reflect stereotypes, i.e., human thinking in group categories, also with regard to gender, cannot initially be blamed on them. After all, stereotypes facilitate information processing (Müller-Spitzer & Lobin, 2022, p. 36). On the other hand, even if dictionaries are intended to be descriptive, they are often perceived as normative authorities. Accordingly, lexicographers have a great responsibility in the design of lexicographical entries, which inevitably convey a certain worldview (Müller-Spitzer & Lobin, 2022, p. 37; Kunkel-Razum, 2012, p. 216). Based on a case study with German newspaper corpora, Müller-Spitzer & Lobin (2022) show a comparatively strong presence of gender stereotypes in this genre, which in turn affect the lexicographical products, e.g., *Duden online* or *elexiko*, especially their automatically generated parts (e.g., typical word combinations in *Duden online*). *Duden online* has been available since 2011. We use it as an example of good practice as it changed and expanded its offering in 2021 with regard to feminine personal and

2.2. Italian-Speaking Area

occupational words.8

Several authors have studied and analyzed gender biases in dictionaries over the last decades.⁹ At the end of the 1980s, the very first Italian contribution to sexism in language was authored by Alma Sabatini (1987). She (1987, p. 28–29) detected numerous asymmetries between meanings and a strongly stereotyped depiction of women.

Lepschy et al. (2001) believed that progressive choices in lexicography are not always suitable. A term might be considered insulting because dictionaries say so rather than because it is really used as insulting (Lepschy et al., 2001, p. 8).

⁵ Eng. 'Popess'.

⁶ Unpublished talk from Kathrin Kunkel-Razum, editor-in-chief of the Duden publishing house on "Arzt, der – männliche Person, die ...; Ärztin, die – weibliche Person, die ...Women and men in current German-language meaning dictionaries", EAAL Conference on 16-17 June 2022, Tallinn.

⁷ "The DUDEN explanatory dictionary as a sequel to a light novel" (translation by the authors).

⁸ Unpublished talk from Kathrin Kunkel-Razum 2022 (see above).

⁹ For a broader overview, please refer to Fusco (2012).

Fusco (2009) shared Sabatini's position on dictionaries' role in society. Her analysis of more than 1,300 lemmas of the *GRADIT* dictionary related to the word *donna*¹⁰ revealed that in many occurrences, women are depicted through stereotypes, especially as to physical and moral attributes as well as to out-of-date traditions. The intellectual value of women was heavily misrepresented (Fusco, 2009, pp. 212–217). In 2019, she analyzed the *GRADIT* dictionary again, finding out that 30 years after the publication of Sabatini's recommendation, many profession names were still not accepted even if documented in dictionaries (e.g., *architetta, poeta, ministra*¹¹). Women were reluctant to employ professional denominations because they did not feel representative of themselves. They still tended to prefer the masculine noun, as it better represented the function they finally came to occupy, according to Fusco (2019, p. 44–46), as well as to Sabatini (1987, p. 30).

Furthermore, Manera & Bazzanella (2006) delved into synonyms, antonyms, and definitions of the *Word Thesaurus* for a selection of terms. Their analysis revealed an evident asymmetry in masculine and feminine representation, severely biased towards women. For instance, synonyms of the word *donna* referred to Christian-practice terms, physical (as a sexual object), moral attitudes, or specific roles (wife, mother). Furthermore, "with regard to the number of meanings, synonyms, and antonyms, the quantitative asymmetry is markedly on the side of the woman" (*Ibid.*, p. 117).

Finally, it is worth mentioning a new feature in the new edition of Treccani's *Dizionario della lingua italiana*, published in 2022, which the publisher describes as "revolutionary" in its advertising of the print version.¹² This includes the lemmatization of feminine forms of nouns and adjectives, which were traditionally only recorded in the masculine form. We will include the online version of this dictionary as a second example of good practice in our study, even if it is not clear from the information available to us how many changes have also found their way into the online edition.

3. E-dictionaries Today

In this section, we will describe by means of two case studies, the current strategies adopted to represent the feminine world in online dictionaries. We will start by analyzing examples of the recent reform that occurred in the *Online-Duden*. Then, we will present analogous examples from *Treccani*, an Italian dictionary breaking new ground in terms of feminine representation, both in its micro- and macrostructure.

Based on the current state of research, the following aspects are in particular considered: a) the presence of female agentives, b) lemmatization, c) descriptions of meaning and examples, and d) editorially vs. automatically created parts in dictionary entries.

¹⁰ Eng 'woman'

¹¹ Feminine form of *architect*, *poet*, and *minister*.

¹² https://emporium.treccani.it/it/il-dizionario-italiano-3012242.html?utm_source=TreccaniIT&utm_medium=ban-ner&utm_campaign=Dizionario_dell_Italiano

When selecting the lemmas for German, we look for examples according to the following linguistic possibilities for gender specification: derivation (*-in* suffix), composition (*-frau*), and substantivization of participles (*-te, -nde*) (see Kotthoff & Nübling, 2018, p. 130; Westweer et al., 2018, p. 375). For Italian, on the other hand, we consider the following: derivation (*-a, -essa, -trice* as suffixes, zero suffix, i.e., identical masculine and feminine forms: *-ante, -ista*¹³) (see Fusco, 2019, pp. 36–37).

3.1 German

For the exemplary analysis, the following nouns were analyzed in Duden online: Anwalt/Anwältin, Autor/Autorin, Architekt/Architektin, Arzt/Ärztin, Bäcker/Bäckerin, Friseur/Friseurin, Feuerwehrmann/Feuerwehrfrau, Abgeordneter/Abgeordnete, Vorsitzender/Vorsitzende, Studierender/Studierende.¹⁴

The descriptions of meaning were adapted for both genders in the course of the revision of *Duden online* by essentially replacing the description of "jemand, der [...]" with "männliche Person, die [...]" on the one hand and introducing "weibliche Person, die [...]"¹⁵ on the other, for example in the lexemes *Arzt* and *Ärztin*. Other examples are comparable, such as *Anwalt* or *Anwältin*, whereby in this case the detour via the explanation of the meaning by means of synonyms *Rechtsanwalt* or *Rechtsanwältin* is to be taken, which in turn is introduced with "Jurist [bzw. Juristin] mit staatlicher Zulassung [...]".¹⁶ Finally, *Jurist* or *Juristin* is described according to the usual pattern of "männliche [or weibliche] Person, die [...]". The same applies to *Architekt* or *Architektin*, who are first introduced as *Fachmann* or *Fachfrau*,¹⁷ while only the entries for the latter follow the familiar pattern. This approach reveals a strategy whereby only one definition of semantically related lemmas follows the usual pattern and tends to be the one in which the *genus proximum* is explained.

In the case of masculine agentives whose feminine form ends in *-in*, as well as substantivized participles, a separate box on the use of the word is inserted, in which the problem of the generic masculine is pointed out, but without using this technical term: "In bestimmten Situationen wird die maskuline Form (z. B. *Arzt, Mieter, Bäcker*) gebraucht, um damit Personen aller Geschlechter zu bezeichnen. Bei dieser Verwendung ist aber sprachlich nicht immer eindeutig, ob nur männliche Personen gemeint sind oder auch andere. Deswegen wird seit einiger Zeit über sprachliche Alternativen diskutiert"¹⁸ (emphasis in original). The use of the substantive present participle for agentives (see e.g., Müller-Spitzer, 2022), which is particularly

¹³ -*tora*, rare, and ignored

¹⁴ Masculine and feminine forms of *lawyer*, *author*, *architect*, *physician*, *baker*, *hairdresser*, *fireman*, *member* of *parliament*, *chairperson*, and *student* [literally 'studying male/female person']

 $^{^{\}rm 15}$ Eng. 'Someone who', 'male person who', 'female person who'.

¹⁶ Eng. '[Male and female] lawyer with a state licence'.

¹⁷ Male and female expert.

¹⁸ "In certain situations, the masculine form (e.g., *doctor, tenant, baker*) is used to refer to people of all genders. However, it is not always clear from a linguistic point of view whether only male persons are meant or others as well. For this reason, linguistic alternatives have been under discussion for some time" (translation by the authors); see e.g., https://www.duden.de/rechtschreibung/Arzt.

controversial in public and linguistic discourse, is provided with an additional box "Besonderer Hinweis"¹⁹, which appears in the article in the form section at the beginning as an announcement, and finally in the meaning section with the note itself: "Als geschlechtsneutrale Bezeichnung oder als Ausweichform für die Doppelnennung *Studentinnen* und *Studenten* setzt sich der Plural *Studierende* immer mehr durch"²⁰ (emphasis in original).

In the lexicographical examples, a quantitative difference between masculine vs feminine agentives is noticeable: In most cases, there are fewer examples for feminine forms. In the entries *Abgeordneter* and *Abgeordnete*, the number is identical, with only the personal designation changing in the examples themselves. In contrast, in the other entries analyzed, the examples are only identical to a small extent. The attribute *behandelnd*²¹, for example, describes both *Arzt* and *Ärztin*, while the attribute *leitender*²² only describes *Arzt*. However, such stereotypical role allocations appear to be rare overall. Variations in meaning, such as *Arzt* or *Ärztin* as an individual vs. as an institution, are not differentiated in the meaning descriptions. They have to be inferred from the examples, e.g., "zum Arzt gehen"²³, for which there is no female equivalent.

Major differences can be recognized in the automatically generated "Typische Verbindungen" for male vs. female agentives. Firstly, it can be seen that in some cases, there are no typical co-occurrences for the female terms at all, e.g., Architektin and Bäckerin. It is noticeable that more co-occurrences between male agentives and other male agentives are shown than with female ones; the opposite is true for female agentives. If we take another closer look at the example of Arzt or Ärztin, we can see that for Arzt there are links to Apotheker, Psychologe, Anwalt, Patient and *Ärztin*, whereas for *Ärztin* there are links to *Psychologin*, *Psychiaterin*, *Juristin*, Rechtsanwältin, Pädagogin, Krankenschwester and Arzt. There are also differences concerning the adjectives: With Arzt we find jung, tätig, jüdisch, praktisch, neu, niedergelassen, diensthabend, tätig²⁴, whereas with Ärztin we find angehend, hübsch, blond, jung, praktisch, niedergelassen, diensthabend.²⁵ The verb combinations are also different: Arzt has verordnen, raten, rufen, verschreiben, sagen, aufsuchen, behandeln, diagnostizieren,²⁶ while Ärztin, by contrast, has arbeiten, empfehlen, fragen, sagen, meinen, antworten, verheiraten, verlieben.27 A number of particularly female gender stereotypes are depicted here, such as hübsch, blond and heiraten.²⁸

¹⁹ Eng. 'Special note'; also for the feminine form.

²⁰ "As a gender-neutral term or as an alternative form for the double mention *female students* and *male students*, the plural "*studying people*" [literal translation] is becoming increasingly popular" (translation by the authors); see e.g., https://www.duden.de/rechtschreibung/Studierende.

²¹ Eng. 'attending [physician]'.

²² Eng. 'head [physician]'.

²³ Eng. 'to go to the doctor'.

²⁴ Eng. 'young', 'active', 'Jewish', 'practical', 'new', 'established', 'serving', 'active'.

²⁵ Eng. 'trainee', 'pretty', 'blonde', 'young', 'practical', 'settled', 'serving'.

²⁶ Eng. 'prescribe', 'advise', 'call', 'prescribe', 'say', 'visit', 'treat', 'diagnose'.

²⁷ Eng. 'work', 'recommend', 'ask', 'say', 'mean', 'answer', 'marry', 'fall in love'.

²⁸ Eng. 'pretty', 'blonde', 'marry'.

GENDER IN ELECTRONIC DICTIONARIES AND TERMINOLOGY DATABASES

3.2 Italian

The following Italian lexemes were selected for the analysis in the online *Treccani* dictionary: *avvocato/avvocata* or *avvocatessa*, *architetto/architetta*, *medico/medica*, *autore/autrice*, *direttore/direttrice*, *cantante*, *professore/professoressa* or *professora.*²⁹ Only some of the female agentives analyzed are included in the dictionary as separate lemmas. It is unclear which criteria play a role here. For example, there is an entry for *medica*; furthermore, there is also an entry for *medichessa* with the indication that this term is used pejoratively or jokingly, while in other cases there is no separate entry for the feminine form at all. Due to the negative connotation that the suffix *-essa* had, especially in the past, the suffix *-a* is sometimes favored for female personal names (see Sabatini, 1987), even if the negative stigmatization no longer tends to be noticeable in usage. Lexems such as *professoressa* or *studentessa* can be used without hesitation (see Fusco, 2019, p. 37).

A look at the grammatical information on the lemmas reveals a rather inconsistent approach: the grammatical information for *avvocato*, for example, refers to the feminine form in brackets as follows: (f.: *-essa* o *-a*). However, this information is not available for all analyzed entries for male persons. Conversely, with some feminine lemmas there is a grammatical indication that it is a derivation of the masculine form, e.g., in *avvocata*: (femine form of *avvocato*). The feminine form *avvocatessa* is not included as a separate lemma in the dictionary. Still in the entry *avvocato* we find a reference to the negative connotation of the term, and the entry *avvocata* contains the information that *avvocatessa* is a synonym that is not in common use (without information on the negative connotation). In addition, *avvocata* does not refer to the professional title in the field of law but only to a meaning in the religious field ("protettrice, interceditrice, attributo della Madonna o di sante"³⁰).

In all entries included in the case study, the entry for the male person has the full scope, while the entry for the female person is very brief. The descriptions of the meanings of the male agentives often follow the person-neutral pattern "chi predispone/professa/ dirige/è ..."³¹, while the definitions of the female agentives are usually introduced with "donna che ...". Due to the brevity of the entries for female persons, they also contain very few examples or none at all.

It is worth mentioning the practice that towards the last part of the dictionary entries for male agentives there is often a separate reference to the female term, e.g., for *avvocato*: "Per indicare una donna che esercita l'avvocatura nell'uso giuridico è stato usato il maschile avvocato, ma sono sempre più frequenti, nell'uso com., i femminili avvocata e avvocatessa, quest'ultimo anche per indicare scherz. la moglie di un avvocato, o una donna che ha la parlantina sciolta, che si accalora nel discorrere e nel sostenere le ragioni proprie o altrui"³² or *medico*: "La forma femm. medichéssa è dell'uso ant., o scherz. e spreg.; nell'uso attuale va diffondersi il femm. medica riferito

²⁹ Eng. 'lawyer', 'architect', 'physician', 'author', 'director', 'singer', 'professor'.

³⁰ "Patroness, intercessor, attribute of the Virgin Mary or saints" (translation by the authors).

³¹ Eng. 'who prepares/provides/manages/is'.

³² See https://www.treccani.it/vocabolario/avvocato/.

a donna che eserciti la professione".³³ The explanations seem to be formulated on an ad hoc basis with reference to the specific case. It remains unclear why the reference is missing in some cases, e.g., with *direttore*.

The extent to which automatically generated parts are included in the dictionary entries cannot be determined, so this aspect is excluded from the analysis.

4. State of the Art in Terminology Science³⁴

Before proceeding, some preliminary considerations are necessary. As stated in TERMCAT (2015, p. 4), lexicography organizes the study of language based on linguistic form: it identifies the canonical form and determines the different meanings that each form can take in real contexts. Therefore, it starts with the word to get to the meaning. Its approach, i.e., the semasiological approach, leads to treating masculine and feminine forms separately based on specific morphological criteria. Instead, terminology science deals with the study of concepts and their designations in specific domains (DIN 2342, 2011, p. 14 in Drewer & Schmitz, 2017, p. 61). The resulting activity, the terminology work, is based on an onomasiological approach. This approach identifies, firstly, the meaning and, secondly, the linguistic or nonlinguistic designation(s)³⁵ assigned to the concept. Hence, in terminology science, it is not the linguistic form of the designation that is decisive but the concept, which, as such, is generally not related to gender (see TERMCAT, 2015, p. 4). In this regard, gender is considered a property, i.e., a qualitative attribute, and not a characteristic, i.e., an intrinsic attribute that forms and delimits concepts (Winter, 2021, p. 29; DIN 2342, 2022, p. 8; ISO 1087, 2019, 3.1.3).

Taking the aforementioned into consideration, how has the terminology world reacted to gender-sensitive and gender-neutral language? To answer this question, in recent years, Bengoechea (2017) analyzed *IATE*, the European Union's inter-institutional terminology database. In particular, Bengoechea investigated how *IATE* dealt with feminine agentives. Her analyses resulted in the conclusion that "very few measures have been taken to implement non-sexist language in private or public terminology databases which include agentives" (Bengoechea, 2017, p. 215).

Drawing on this investigation, between 2020 and 2021, we also explored some terminology resources such as *IATE*, *TERMDAT*, *LinguaPC*, *TERMCAT*, *Euskalterm*, *FAOTERM*, *Termium Plus*®, and *Le Grand dictionnaire terminologique (GDT)*, focusing on feminine agentives. The trend identified by Bengoechea seemed to be still current: in some databases, feminine agentives were not entered systematically if they were entered at all (Ralli & Evers, in review). For example, in IATE there

³³ See https://www.treccani.it/vocabolario/medico3/.

 $^{^{34}}$ In this section (4) and in the following sections on terminology databases (5.1 and 5.2), the agentives are translated into English exclusively when they are present in the respective databases. For the same reason, their definitions (5.1 and 5.2) are only given in the language provided in the respective databases.

³⁵ A concept can also be represented by definitions, formulas, symbols, graphics, etc.

was a matching entry³⁶ for the query *infermiera*³⁷ with the corresponding form in German "Gesundheits- und Krankenpflegerin", but there was no matching entry for *sindaca*³⁸. Other databases tried to treat systematically women and men on an equal footing by making "women linguistically visible besides men" (Motschenbacher, 2014 in Bengoechea, 2017, p. 206). That means feminine and masculine agentives were recorded in the same entry as independent items. This allows searching for female professionals, job titles, etc. via query mask.

This general analysis was the starting point for more in-depth research on our part. Different aspects needed to be closely analyzed:

- a. the presence of feminine agentives;
- b. how feminine agentives are recorded: whether gender is considered a property or a characteristic leads to different representations of the feminine agentives in terminology databases (Ralli & Evers, in review);
- c. to what extent feminine agentives are considered, i.e., only when they occur
 1) as simple terms, e.g., *Richterin*³⁹ and 2) as complex terms, e.g., *nebenamtliche Richterin*⁴⁰, or also when they are part of complex terms, e.g., *auf Geheiss der Richterin*⁴¹;
- d. how the sexes are treated graphically (see Bengoechea, 2017, p. 217);
- e. how agentives are defined, i.e., whether a gender-sensitive and a genderneutral language is used or not. Regarding this point, a clarification is necessary (Ralli & Evers, in review): in terminology work, concepts are generally defined by *genus proximum* and *differentiae specifica* (Magris, 1998, p. 41). Thus, the concept is described according to its intension, by stating its (closest) generic concept and specifying the delimiting characteristics (see Arntz et al., 2014, pp.63 ff.; ISO 704, 2022, 6.2). Our analysis focused on how and to what extent terminology databases can be gender-sensitive and gender-neutral, particularly when defining hyponyms.

To achieve these goals, we analyzed three well-known multilingual databases from Switzerland and Catalonia: *LinguaPC, TERMDAT*, and *TERMCAT*. We chose these databases because they proved to be good practice in our first investigation.

 $^{^{36}\} https://iate.europa.eu/entry/slideshow/1631775442783/1430984/it-de-en,\ consulted\ on\ 15th\ September\ 2021.$

³⁷ Feminine form of *nurse*.

³⁸ Feminine form of *mayor*.

³⁹ Feminine form of *judge*.

⁴⁰ Feminine form of *substitute judge*.

⁴¹ "By order of the judge", where *Richterin* is the feminine form of *judge*.

5. Terminology Database Today

5.1 LinguaPC and TERMDAT

LinguaPC and *TERMDAT* are Swiss public terminology databases: the first is from Canton Bern and contains legal and administrative terms in German and French; the second is from the Federal Chancellery and includes the terminology of Swiss law, public administration, and the public sector in German, French, Italian, Romansh⁴², and, sometimes, in English⁴³. However, for space reasons, we focus exclusively on the German part of the terminological entries.

In both databases, German feminine and masculine agentives are treated as synonyms. For example, in *LinguaPC* the entry *Anwalt*⁴⁴ contains the corresponding feminine form *Anwältin*. Its synonyms *Rechtsanwalt* and *Fürsprecher* have the corresponding feminine forms *Rechtsanwältin* and *Fürsprecherin* too. The same applies in *TERMDAT*, where under *Anwalt*, *Anwältin*, *Fürsprecher*, and *Fürsprecherin* other synonyms are recorded: *Advokat* and *Advokatin*⁴⁵. Synonymity implies that all these agentives are lemmatized as separate items in the same terminological entry they belong to. That means in the query mask, you can start your search by entering the masculine or feminine form of an agentive, getting the same search results. Taking the above-mentioned example, you can search for both *Anwalt* and *Anwältin* or *Fürsprecher* and *Fürsprecherin*.

Masculine and feminine forms are also present in complex terms such as eingetragener Anwalt and eingetragene Anwältin. Also, in this case, the same rules apply as for the simple terms. The two investigated databases follow different strategies in complex terms, where an agentive is an object. LinguaPC tends to record both masculine and feminine agentives, such as Patent eines bernischen Kindergärtners⁴⁶ (m.) and Patent einer bernischen Kindergärtnerin (f.). However, this database does not contain enough complex terms of this type to confirm this assumption. On the contrary, *TERMDAT* tends to use the masculine agentive. For example, this is the case of Ausschluss eines Miteigentümers or Abberufung eines Verwalters where no feminine form (Auschluss einer Miteigentümerin, Abberufung einer Verwalterin) is given in the terminological entry. The decision to omit the feminine form could have a methodological reason: in general, this database adopts a descriptive approach and documents designations as they are used in contexts. The presence of only masculine agentives may indicate a lack of authoritative and reliable sources attesting to the existence and use of the corresponding feminine agentives.

⁴² Romansh is a Rhaeto-Romance language spoken in the in the Swiss canton of the Grisons (Graubünden).

⁴³ https://www.bk.admin.ch/bk/en/home/dokumentation/languages/termdat.html (retrieved June, 12, 2024).

⁴⁴ Masculine form of *lawyer*.

⁴⁵ All the examples in this section are from *LinguaPC* (https://www.linguapc.apps.be.ch) and/or *TERMDAT* (https://www.termdat.bk.admin.ch), consulted until June 12, 2024. Wherever possible, the English translations of the terms used as examples for *LinguaPC* are sourced from *TERMDAT*.

⁴⁶ In this section, boldface has been added by the authors.

With regards to definitions, gender-neutral expressions (i.e., impersonal expressions, collective nouns) are generally used when a superordinate concept is defined. For example, LinguaPC defines Anwalt and Anwältin starting with the gender-neutral term person: "Person, die vom Obergericht das Anwaltspatent erhalten hat". On the contrary, the definition of the hyponyms amtlich bestellter Anwalt or eingetragener Anwalt begins with a masculine designation, which represents their immediate superordinate concept, i.e., Verteidiger⁴⁷ and Anwalt respectively. The same pattern can be observed in other entries. For example, the definition of *Richter* and *Richterin*⁴⁸ is formulated as follows: "Vom Grossen Rat für eine Amtsdauer von sechs Jahren gewähltes Mitglied einer kantonalen Gerichtsbehörde", where "Mitglied" corresponds to the gender-neutral term *member*. In the case of hyponyms like *ordentlicher Richter*⁴⁹ or Einzelrichter⁵⁰, LinguaPC defines them beginning with the masculine agentive *Richter* as its closest superordinate concept. *TERMDAT* applies the same approach: many agentives are defined using the masculine form or gender-neutral words.⁵¹ In addition, besides impersonal expressions or nouns, combined forms are also used here. For instance, TERMDAT defines Bundesrichter and Bundesrichterin⁵² as "Richter/ Richterin am Schweizerischen Bundesgericht". Considering this analysis, however not exhaustive, we can claim that much is being done in both databases to apply the principle of gender equality in terminology definitions, wherever possible.

5.2 TERMCAT

TERMCAT is a terminology consortium in charge of normalizing terminology in Catalan. Its mission includes, inter alia, the maintenance, and update of a Catalan terminology database with equivalent terms in other languages.⁵³

In its terminology database, masculine and feminine agentives in Catalan and Spanish are recorded together within the same field and separated by a glyph ("|) as shown in the following example in Catalan⁵⁴:

1. advocat | advocada, advocadessa nm, f

This type of lemmatization violates two terminological principles (Ralli & Evers, in review): the first concerns term autonomy, according to which all terms in a terminological entry "are considered independent sub-units and can be described using

⁴⁷ Masculine form of *defence lawyer*.

⁴⁸ Masculine and feminine form of *judge*.

⁴⁹ Masculine form of *ordinary judge*.

⁵⁰ Masculine form of *single judge*.

⁵¹ See also the agentives *Richter* and *Richterin, unfähiger Richter* and *unfähige Richterin, Anwalt* and *Anwältin, Tierschutzanwalt* and *Tierschutzanwältin, Bundesanwalt* and *Bundeanwältin*, just to mention a few (retrieved June, 12, 2024).

⁵² Masculine and feminine form of *federal judge*.

⁵³ https://llengua.gencat.cat/en/direccio_general_politica_linguistica/02_organismes_vinculats/01_centre_de_terminologia_termcat/ (retrieved June, 13, 2024)

⁵⁴ All the examples of this section are from *TERMCAT*'s terminology database CERCATERM (https://www.termcat. cat/ca/cercaterm/advocata?type=basic&thematic_area=&language=) consulted before 13th June 2024.

the same set of data categories" (ISO 26162-1, 2019, 3.2.14); the second relates to data elementarity, according to which "a data field contains only one data element" (ISO 26162-1, 2019, 3.2.16). Nevertheless, it puts the masculine and feminine agentives at the same level, thus ensuring graphically equal treatment. Furthermore, this does not prevent searching for both forms via query mask, obtaining the same search results.

The same strategy applies both for simple terms and complex terms such as defense lawyer: *advocat defensor* (m.) and *advocata defensora* (f.) in Catalan or *abogado defensor* (m.) and *abogada defensora* (f.) in Spanish in procedural law. Regarding complex terms, no feminine form is provided where an agentive is an object.⁵⁵ This is the case of *petició d'advocat d'ofici*⁵⁶ in Catalan and *petición de abogado de oficio* in Spanish. The reason behind this choice may be the same as that described in Section 5.1. Finally, in the case of epicene designations, i.e., designations "that may refer to a male or a female"⁵⁷, like the Catalan word *pilot*⁵⁸, there is a double attribution of grammar, and the designation is entered once to avoid redundancy (TERMCAT, 2015, p. 3):

2. pilot, n m f⁵⁹.

Concerning definitions, these are only in Catalan as this is the pivot language of the database. As in the case of *LinguaPC* and *TERMDAT*, also *TERMCAT* tends to use gender-neutral terms for defining superordinate concepts. For example, the definition of *advocat*, *advocada*, and *advocadessa* begins with different gender-neutral terms according to the subject field the agentives belong to:

- 3. **"Persona** que assessora sobre qüestions de dret i defensa les causes civils, criminals o laborals d'altri." (procedural law);
- 4. "Professional que es dedica a intervenir en causes o processos judicials i administratius, i a donar orientació i assessorament en matèria legal, tant a persones particulars com a professionals de l'àmbit de l'atenció a les persones" (welfare).

As is evident from these examples, also the delimiting characteristics are described with a gender-neutral language, too.

Also, in the case of subordinate concepts, we can observe similarities with the two Swiss databases. Their definitions usually begin with the designation of the respective superordinate concept, as in the entry *advocat matrimonialista* and *advocada matrimonialista*: "Advocat especialista en processos de separació i divorci", where *advocat especialista* is the hyperonym of *advocat matrimonialista* in the masculine

⁵⁵ See as examples also sol·licitud de designació d'advocat (ca), solicitud de designación de abogado (es), "col·legi d'advocats" (ca), colegio de abogados (es) (retrieved June, 12, 2024).

 $^{^{\}rm 56}$ In this Section, boldface has been added by the authors.

⁵⁷ https://www.collinsdictionary.com/de/worterbuch/englisch/epicene (retrieved June, 12, 2024)

⁵⁸ Eng. 'pilot'.

⁵⁹ https://www.termcat.cat/ca/cercaterm/pilot?type=advanced&thematic_area=&language=&condition=match&fields=denominacio&category=&hierarchy=&page=1 (retrieved June, 12, 2024).

form. Sometimes subordinate concepts are defined using gender-neutral designations or both masculine and feminine forms. This is the case of *advocat defensor* and *advocada defensora*⁶⁰:

5. "**Advocat o advocada** que defensa una persona que litiga, especialment en un procés penal" (procedural law).

At the time of Bengochea's investigations (2017, p. 215) *TERMCAT* incorporated female agentives only for Catalan. Our analysis shows that now the Catalan database incorporates such agentives for Spanish, too. Like in *LinguaPC* and *TERMDAT*, respect for gender equality goes beyond the lemmatization of feminine agentives to the drafting of definitions in gender-neutral and gender-sensitive language. A basis for ensuring linguistic visibility and equality has been laid down.

6. Conclusion and Outlook

In German- and Italian-speaking countries respectively, *Duden* and *Treccani* are regarded as examples of good practice for publishers that have adopted new approaches to more gender-sensitive language use, and they did so in roughly the same period, i.e., 2021/2022. Therefore, the online versions of their dictionaries formed the basis for our case studies. Both pursue strategies to make female agentives more visible in their products. Both dictionaries have their own explicit information on the changing use of male and female agentives. In *Duden online*, this information, formulated according to a standardized pattern, can be found in separate info boxes on the website of each dictionary entry on male agentives. In *Treccani* online, the information is part of the dictionary entry (which closely follows the format familiar from print dictionaries) on masculine agentives.

In *Duden online*, female agentives are consistently lemmatized, followed by a complete dictionary entry. *Treccani* online is less consistent in this respect, so there is no separate entry for all feminine forms. It is noticeable in both dictionaries that the entries for female agentives are less extensive than those for male personal names, although the difference in *Treccani* online is more significant. There, women are relatively marginally visible, at least in the context of the case study.

Duden online uses a standardized definition practice, which usually refers to either a "männliche Person, die …" or a "weibliche Person, die …". In *Treccani* online, this strategy is applied, albeit not consistently, to female agentives.

Some tendencies to perpetuate stereotypical role models are still recognizable. In *Duden online*, this applies in particular to the automatically generated typical word combinations. What is striking in *Treccani* online is the fact that a pejorative register marker often accompanies female personal names This may be due to a peculiarity of the Italian language, but the dictionary certainly reflects this linguistic reality.

⁶⁰ See also professor de formació viària and professora de formació viària in the frame of road safety or ajudant del jutge àrbitre and ajudanta del jutge àrbitre (retrieved July, 31, 2024).

In the terminology field, *LinguaPC*, *TERMDAT*, and *TERMCAT* are among the databases that better apply gender equality in their data. From their representation, we could observe different underlying methodological choices. Considering our investigation, a common methodological approach is lacking, which would be relevant to facilitating data interchange.

We believe that many aspects need careful consideration, among them (Chiocchetti & Ralli, 2024; Ralli & Evers, in review):

- a. lemmatization of feminine agentives: we can have different types of representations depending on whether the feminine agentive is considered a property or a characteristic. If gender is considered a property, the feminine agentive is represented with the masculine agentive in the same entry. In this case, the following two scenarios are envisaged: the first, the feminine form is recorded as a synonym of the masculine agentive (see Section 5.1); the second, the feminine form is recorded together with the masculine agentive(s) within the same field (see Section 5.2). Instead, if gender is considered a characteristic, the feminine agentive is treated as a concept, thus in a dedicated entry and separated from the masculine agentive. In this case, a semasiological approach would be followed;
- b. term type and degree of syntagmatic complexity: feminine and masculine agentives can occur as simple terms (e.g., *Anwalt/Anwältin*), as complex terms (e.g., *eingetragener Anwalt/eingetrage Anwältin*), or as an object of a complex term with agentives as objects (e.g., *Patent eines bernischen Kindergärtners/ Patent einer bernischen Kindergärtnerin*);
- c. drafting of definitions: the generic use of masculine agentives to refer to the closest generic concept (see Sections 5.1 and 5.2) can be avoided by combining several strategies,⁶¹ wherever possible. Some of them are also applied by *LinguaPC*, *TERMDAT*, and *TERMCAT*. However, as Ralli and Evers (in review) state: "the relationship between the hypernym and hyponym is semantic, i.e., linguistic, rather than conceptual because in this case the gender is treated as a property and not as a characteristic."

The Council for German-Language Terminology, RaDT⁶², is working on all these aspects, among other things. A publication in this regard is planned soon.

Of course, we cannot consider our investigations exhaustive: more in-depth research is needed, also with regard to other languages and the impact of such representation on society. However, our analysis shows that it is high time for lexicographers and terminologists to deal with this topic systematically, studying shared standards for representing men and women more fairly and equally in lexical resources.

⁶¹ Such strategies should be applied according to each language's grammatical typology (European Parliament 2018, p. 5).

⁶² Rat für deutschsprachige Terminologie, http://radt.org/

GENDER IN ELECTRONIC DICTIONARIES AND TERMINOLOGY DATABASES

References

Arntz, R., Heribert, P., & Schmitz, K-D. (2014). Einführung in die Terminologiearbeit. 7. OLMS.

Bazzanella, C. (2009). Stereotipi e categorizzazioni del femminile/maschile. In S. Regazzoni (Ed.), *Mi fai male... Atti del Convegno Venezia, Auditoriume Santa Margherita 18-19-20 novembre 2008*. Cafoscarina.

Bengoechea, M. (2017). Agentives for women, a gap still to fill: dismissing non-sexist language policies in terminological resources. In P. Faini (Ed.), *Terminological approaches in the European context* (pp. 200–221). Cambridge Scholars Publishing.

Calvo Ramos, L. (1998). Algunos rasgos sexistas en el Diccionario de la Real Academia Española de la lengua. *Revista Llengua i Dret, 30*, 47–62.

Cavagnoli, S., & Dragotto, F. (2021). Sessismo. Mondadori Università.

Chiocchetti E, & Ralli, N. (in review). Comunicazione istituzionale inclusiva in Alto Adige: esperienze e questioni aperte. *Convegno Internazionale di Linguistica e Glottodidattica Italiana CILGI5 2023*. March 23-25, 2023, Salerno.

De Bernardis, I. (2007). Reviewed Work(s): Piccolo dizionario dell'inuguaglianza femminile by Alice Ceresa and Jacqueline Risset. *Belfagor*, *62*(5), 616–619.

DIN 2342 (2022). Terminologiewissenschaft und Terminologiearbeit-Begriffe. Beuth.

Drewer, P., & Schmitz, K.-D. (2017). Terminologiemanagement. Grundlagen – Methoden – Werkzeuge (Kommunikation und Medienmanagement). Springer Vieweg.

European Parliament (2018). *Gender-neutral language in the European Parliament*. Retrieved June 15, 2024, from https://www.europarl.europa.eu/cmsdata/151780/GNL_Guidelines_EN.pdf

Euskalterm: *Banco Terminológico Público Vasco*. Retrieved September 18, 2020, from https://www.euskadi.eus/gobierno-vasco/contenidos/informacion/euskalterm/es_7553/euskalterm.html

Evers, E. (2022). Geschlechtersensible Sprache – wie bringe Ich es meiner Terminologiedatenbank bei? *edition 2*/22, 13–20.

FAOTERM: *FAO Terminology Portal*. Retrieved September 18, 2020, from https://www.fao.org/faoterm/en/

Fuertes-Olivera, P. A., & Tarp, S. (2022). Critical Lexicography at Work: Reflections and Proposals for Eliminating Gender Bias in General Dictionaries of Spanish. *Lexikos*, *32*(2), 105–132.

Fusco, F. (2009). Stereotipo e genere: il punto di vista della lessicografia. *Linguistica*, 49(1), 205–225. doi:10.4312/linguistica.49.1.205-225

Fusco, F. (2019). Il genere femminile tra norma e uso nella lingua italiana: qualche riflessione. In S. Adamo, G. Zanfabro, & E. Tigani Sava (Eds.), *Non esiste solo il maschile. Teorie e pratiche per un linguaggio non discriminatorio da un punto di vista di genere* (pp. 27–49). EUT. Edizioni Università di Trieste. Graham, A. (1975). The making of a non-sexist dictionary. In B. Thorne, & N. Henley (Eds.), *Language and Sex: Difference and Dominance* (pp. 57–63). Newbury House Publishers.

Hampares, K. J. (1976). Sexism in Spanish Lexicography? *Hispania*, 59(1), 100-109. doi:10.2307/339381

Hanks, P. (2012). The Corpus Revolution in Lexicography. *International Journal of Lexicography*, 25(4), 398–436. https://doi.org/10.1093/ijl/ecs026

IATE: *Interactive Terminology for Europe*. Retrieved June 12, 2024, from https://iate.europa.eu/ ISO 1087 (2019). *Terminology work and terminology science—Vocabulary*. ISO.

ISO 704 (2022). Terminology work—Principles and methods. ISO.

Kotthoff, H., Nübling, D., & unter Mitarbeit von Schmidt, C. (2018). *Genderlinguistik: eine Einführung in Sprache, Gespräch und Geschlecht.* Narr Francke Attempto.

Le Grand dictionnaire terminologique (GDT). Retrieved September 18, 2020, from https:// vitrinelinguistique.oqlf.gouv.qc.ca/ (last consulted on 18th September 2020)

Lepschy, A. L., Lepschy, G., & Sanson, H. (2001). Lingua italiana e femminile. *Quaderns d'Italià* 6, 9. doi:10.5565/rev/qdi.51

Lepschy, G. (1987). Sexism and the Italian language. *The Italianist*, *7*(1), 158–169. doi:10.1179/ ita.1987.7.1.158

Lingua-PC: *Terminologiedatenbank des Kantons Bern*. Retrieved June 12, 2024, from https://www.linguapc.apps.be.ch/

Magris, M. (1998). La definizione in terminologia e nella traduzione specialistica. Rivista internazionale di tecnica della traduzione. International Journal of Translation, 3, 37–63.

Manera, M., & Bazzanella, C. (2006). Gender on-line in the Italian Word Thesaurus. In M. Hellinger, & H. Bußmann (Eds.), *Gender, language and new literacy* (pp. 107–122). Continuum.

Müller-Spitzer, C., & Lobin, H. (2022). Leben, lieben, leiden: Geschlechterstereotype in Wörterbüchern, Einfluss der Korpusgrundlage und Abbild der sprachlichen, Wirklichkeit. In G. Diewald, & D. Nübling (Eds.), *Genus – Sexus – Gender* (pp. 33–649).

Müller-Spitzer, C. (2022). Zumutung, Herausforderung, Notwendigkeit? Zum Stand der Forschung zu geschlechtergerechter Sprache. *Geschlechtergerechte Sprache*, 23–29.

Nübling, D. (2009). Zur lexikografischen Inszenierung von Geschlecht. Ein Streifzug durch die Einträge von Frau und Mann in neueren Wörterbüchern 37(3), 593–633. https://doi.org/doi:10.1515/ZGL.2009.037

Porsch, P. (2004). Frau im Wörterbuch – Das DUDEN-Universalwörterbuch 2003 als Fortsetzung eines Trivialromans. In U. Fix, M. Schröder, & H. Wellmann (Eds.), *Zwischen Lexikon und Text – lexikalische, stilistische und textlinguistische Aspekte* (pp. 358–365). Verlag der Sächsischen Akademie der Wissenschaften zu Leipzig.

GENDER IN ELECTRONIC DICTIONARIES AND TERMINOLOGY DATABASES

Ralli N., & Evers, E. (in review). To gender or not to gender, that is the question: gender-inclusive language in the legal context. EAFT Summit 2023.

Sabatini, A. (1987): *Il sessismo nella lingua italiana*. Commissione Nazionale per la parità e le pari opportunità tra uomo e donna.

TERMCAT (2015): Denominació: Representació del femení (particularitats). TERMCAT.

TERMCAT: Cercaterm. Retrieved June 12, 2024, from https://www.termcat.cat/ca

TERMDAT: *The Federal Administration's terminology database*. Retrieved June 12, 2024, from https://www.bk.admin.ch/bk/en/home/dokumentation/languages/termdat.html

Termium Plus®: *the Government of Canada's terminology and linguistic data bank*. Retrieved June 12, 2024, from https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng

Thüne, E.-M., & Leonardi, S. (2006). The German Word Thesaurus and socio-cultural models. In E.-M. Thüne, S. Leonardi, & C. Bazzanella (Eds.), *Gender, Language and New Literacy: a Multilingual Analysis.* (pp. 79–94). Continuum.

Westveer, T., Sleeman, P., & Aboh, E. O. (2018). Discriminating dictionaries? feminine forms of profession nouns in dictionaries of French and German. *International Journal of Lexicography 31*(4), 371–383.

Whitcut, J. (1984). Sexism in dictionaries. In R. R. K. Hartmann (Ed.), *LEXeter '83: proceedings* (pp. 141–144). De Gruyter. doi:10.1515/9783111593166-020

Winter T. (2021). Das Gendern. Terminologisch betrachtet ein Missverständnis mit Skandalpotential. *edition 2*/21, 29-30.

Acknowledgements

We kindly thank experts from the Central Language Services, Terminology Section of the Swiss Federal Chancellery, and the Central Service of Terminology of the State Chancellery of the Canton of Bern for their valuable input. We would also like to thank our colleague Flavia De Camillis for working with us on the study.

Contact information

Andrea Abel

Free University of Bozen-Bolzano – Faculty of Education / Eurac Research – Institute for Applied Linguistics andrea.abel@unibz.it andrea.abel@eurac.edu

Natascia Ralli

Eurac Research – Institute for Applied Linguistics natascia.ralli@eurac.edu