Women, Who Climb – A Corpus Linguistic Tour Description with Potential Danger Zones

Claudia Posch

Department of Linguistics, University of Innsbruck

Abstract: This paper explores novel ways in which Corpus Linguistics, Critical Discourse Studies (DHA – the discourse-historical approach), and Feminist Linguistics can be fruitfully combined and thus contribute to a new awareness and knowledge concerning language and gender in discourse. For the paper a large heritage corpus of texts on mountaineering (*Alpenwort – Corpus of the Austrian Alpine Club Journal*) is explored, which constitutes a type of discourse that has yet rarely been studied in the abovementioned fields. The corpus was used to analyse gendered nomination with the following questions in mind: What frequently occurring patterns and typical ways of producing gendered nomination can be observed? How are the binary categories 'female' and 'male' linguistically produced and reinforced in these texts? These questions are answered using standard corpus linguistic methods: for example, gendered key items are determined via historical keyness-analysis; the concept of c-collocates and the method UFA (Usage Fluctuation Analysis) are used to trace how the meaning of such key items develops over time. These procedures reveal salient patterns of gender referencing that are then interpreted with the background of the DHA framework and can thus contribute to the investigation of gender bias in sports/leisure discourse.
Keywords: Digital Linguistics, diachronic analysis, mountaineering discourses, c-collocation, key items

The history of mountaineering has long been a history of men: it was written by men and it is mostly about them (Runggaldier 2011: 9). It is thus interesting to see if and how women are represented in mountaineering discourse and if there are historical changes in the ways women are conceptualised in this discourse strand. Based on the history

of mountaineering we might expect women to be discursively underrepresented[1] in texts about this topic. Furthermore, in terms of the form of how they are represented, we may expect aspects of objectification, trivialisation, and stereotyping. To study the representation of women in the dataset, frequently occurring patterns of reference and typical ways of linguistically creating the category 'woman' were analysed.

Related work in corpus and (feminist) discourse studies

This study follows an approach, Digital Linguistics (DL), which links Corpus Linguistics and the DHA (Discourse Historical Approach) framework as proposed in Baker et al. (2008) and Reisigl and Wodak (2016). The benefits of combining these frameworks for uncovering ideologies and evidence of disadvantages have been theorised in a large body of research (for example in Baker et al. 2008; Baker, McEnery 2015b; Gabrielatos, Baker 2008a; Hardt-Mautner 1995; Partington 2004; Partington 2015; Posch 2022; Stefanowitsch 2020; Taylor, Marchi 2018).[2] Two important points shall be summarised here briefly: analysing texts with corpus statistical tools allows us to study larger amounts of text than in the traditional DHA. This makes what is analysed reproducible and enables the researcher to make data-driven, reproducible selections of what to analyse. This data analysis results in further data that need to be interpreted, which can only be done meaningfully with a knowledge foundation in hermeneutic approaches, such as the DHA.

There are numerous previous studies that investigated gendered language in different general and themed text corpora. For example, one early study explored the frequencies of lexical items denoting women/men or females/males (Kjellmer 1986) and found more terms overall for men than women in the data used (Brown and LOB corpus). In a study of the items 'man'/'woman' and their verbal collocates Kilgarriff and Tugwell (2002) found significant differences between the verbs these two items collocate with. In a more recent corpus-based comparison of gendered terms in four diachronic corpora of British English, Baker (2010: 145) discovered that a male bias is verifiable in quantitative terms but also that the usage of male terms seems to be decreasing (notably *he* and *Mr*). In another combinatory corpus and discourse study on the representation of gender and agency using the Harry Potter series as a corpus, Hunt was able to show that in fact the representation of agency in these books is clearly gendered and female characters are generally represented as physically deficient

---

[1] Non-conforming gender identities will be even more under-represented.

[2] Related research endeavors may use different labels for similar approaches, such as Corpus-Assisted Discourse Studies (Partington 2013), corpus-based CDA (Gabrielatos, Baker 2008b), Corpora and Discourse Studies (Baker, McEnery 2015a), Corpus Pragmatics (Scharloth, Bubenhofer 2012) and recently Digital Linguistics (Posch 2022).

(Hunt 2015). Furthermore, in an extensive investigation using the German COSMASII corpus tool, Lautenschläger (2016) looked at German press texts and how they (re)produce images of gender and bodies linguistically. This study found that the concept of masculinity is referred to much more frequently overall, and that femininity is often represented focusing on the topic of gender differences and disparities (and not so often in more general texts). An analysis of linguistic constructions of gender in US sports reportage (basketball) revealed that reporting on the women's basketball league is always marked and thus contributes to the 'othering' of women and women athletes in sports news reporting (Aull, Brown 2013). Ismail (2016), investigating gendered keywords in Malaysian sports news discourse, moreover found the gender marking of sports competitions and associations and that female athletes are often associated with emotions. Last but not least, in their corpus-assisted discourse study of gender representations in the British press around the London Olympics, Jaworska and Hunt (2017) found an increase in balanced gender representations in the press around the Olympics. However, they also said there was little carry-through in the years after the Olympics and there exists a strong intersection with between gender representation and nationalism.

Positionality statement

This positionality statement is intended to self-critically position this work within qualitative and quantitative research as proposed in McCuaig et al. (2019) and Secules et al. (2021). As the author of this study, I am aware of my privileged social position with a biography of a white cis-woman[3] growing up in Europe. By stating this explicitly I would like to contribute to deconstructing the effect of social positioning (Tudor 2011) and would like to define the overall approach of this work as anti-sexist.

Data and methods

The question of gender representation in mountaineering was studied using the zav_20200416 version of the domain-specific Alpenwort corpus (Posch, Rampl 2018; Rampl, Posch 2019), a 19.9-million-word corpus covering the entire *Austrian Alpine Club Journal* (AACJ) from 1870 to 2010. The AACJ is a relevant source for mountaineering discourses because it contains a variety of text types in each volume, from scientific articles, which tend to play a larger role in the early publications, to numerous ascent and expedition reports, literary texts, and smaller text types. The

---

[3] For an explanation of the term 'cis', see Tudor and Hornscheidt (2011: 181) or https://www.merriam-webster.com/dictionary/cisgender.

journal also covers a range of topics relevant to readers interested in mountains in general and mountaineering in particular (Posch, Rampl 2020). The corpus was designed for discourse-related research questions, especially on gender and nationality, along the lines of Baker and McEnery (2015a), Mautner (2016), Rheindorf and Wodak (2019a), and Rheindorf and Wodak (2019b). Therefore, it can be considered 'maximally representative' (McEnery, Brookes 2022: 96) of the context of interest. Alpenwort is fully part-of-speech annotated and continually improved.[4] The data are diachronically segmented into sub-corpora, one for every decade of the *AACJ's* span of publication. The corpus is freely available and can be queried via the online tool CQPweb (Hardie 2012) at http://sprawi-cqp.uibk.ac.at.[5] CQPweb is a web-based corpus analysis system with an extensive search syntax and statistical analysis options such as concordance lines, collocation analysis, keyness, and frequency lists (see below for an explanation of the relevant statistical analysis functions used in this paper below).

Concerning the topics or genres of DL data, Lischinsky (2018: 63) critically remarked that so far there has been a focus on certain 'high flying' genres like news reporting, political speeches, etc., and called for a greater focus on 'under-researched content, both in terms of topics and text types' (Lischinsky 2018: 17). The present research pursues this goal by studying mountaineering, which has hitherto rarely been investigated in linguistics, with the exception of work by Bubenhofer and Scheurer (2014) and Bubenhofer and Schröter (2012). Mountaineering is a sport without an audience and thus relies heavily on communicative events that construct what constitutes a mountaineering event. Climbs and tours, for example, need to be verbally (and graphically) accounted for so that they are deemed valid within the community. Alpinist discourses can as a result be found over a long period of time in a multitude of text types, such as biographies, climbing reports, documentaries, films, interviews, guide literature, route descriptions, and even fiction. In terms of their content and topoi these texts therefore constitute a certain discourse genre or even a rhetoric in the sense that they represent a specific socially accepted way of using language in connection with a particular type of social activity (Wodak 2008: 15). The language of mountaineering discourses thus offers a glimpse into a public discourse that is removed from the direct political sphere but is nonetheless highly political (Achrainer, Kaiser, Trojer 2011: 9). Mountaineering is interesting for a DL study because of (a) the abundance of text sources and (b) its relevance as a domain of public discourse.

Furthermore, the connection to gender is important in mountaineering discourses,

---

[4] For a detailed overview of the Alpenwort corpus, its content, the process of corpus design and creation, and alternative versions, see Posch and Rampl (2020).

[5] For a test log-in and username, see the instructions on the start page.

as such discourses represent a way of climbers relating to other climbers the very concept of a 'climbing identity' and how to be a climber (Rak 2007: 112). To be accepted as a member of the in-group of mountaineers, certain discursive structures and patterns have to be followed – for example, there is a specific style one adheres to and there are specific topoi that are repeatedly activated (e.g. war metaphors, solitude, competition of nations, etc.). The discourses around mountaineering are, for example, closely related to Western identity constructions and are therefore well suited for a study of the identity category 'gender'. Bubenhofer and Schröter (2012: 265) argue that what we find in mountaineering discourses are conceptualisations, beliefs, evaluations, and behaviours that are shared by mountaineers (e.g. climbers, alpinists) and thus the traces of what forms the mountaineering mentality. And Lutz (2002) describes mountaineering as a form of primarily masculine encounter with the body and nature, which is 'hypermasculine'.[6]

The methods used in this work adhere to the Digital Linguistics framework and have been described in detail in previous work (Ismail 2016; Posch 2022; Posch, Rampl 2020; Rampl, Posch 2022). For this paper, the concepts of 'keyness' as developed by (Gabrielatos 2018) and 'historical key-items' or 'key-(key) items' (Krendel, McGlashan, Koller 2021) are used as a starting point for analysis. Keyness has been described as referring to the aboutness of texts (Gabrielatos 2018: 226) and *key items* are words whose frequency in a corpus is disproportionately high compared to their occurrence in a reference corpus. This is usually calculated using an effect-size statistical metric, such as the LogRatio algorithm (Hardie 2014), which is implemented as a functionality of the CQPweb keywording tool.[7] The output of the tool is a list of words that are statistically more salient in the corpus of interest than in the comparison corpus. Historical key-items or key-key items are consequently items that are key items in more than one unit of time, in the case of Alpenwort in more than one decade. The more segments a key-key item appears in, the more 'key-key' it is.

Departing from the most salient key-key item regarding gender, further distributional analysis was conducted. Here, the concept of relative or normalised frequencies is important in order to make frequencies comparable (because not all texts have the same length). Relative frequency is calculated as described in Figure 1 below. The CQPweb corpus analysis tool uses 1 million as the basis for normalisation; hence words are compared according to their frequency per million words (fpmw).

---

[6]   Other identity concepts have connotations as well, such as the 'nation' and 'soldiery', and the inter-discursive ties of mountaineering with colonialism are quite apparent (Purtschert 2019).

[7]   There are numerous algorithms that are used for the purpose of calculating key items and collocations and there is an ongoing discussion about which statistical measure is preferable. For an overview of the discussion, see Gabrielatos (2018) as well as the excellent power point slides by Gabrielatos and Marchi (2012). For an excellent explanation of the statistics for non-mathematicians see Brezina (2018).

Figure 1: Formula for calculating normalised frequency from Brezina (2018: 43)

$$\text{relative frequency} = \frac{\text{absolute frequency}}{\text{number of tokens in corpus}} \times \text{basis for normalization}$$

Source: Author.

Keyness analysis alone cannot be used to study the contexts in which words appear, as it only tells us which items are statistically salient. To contextualise key items, collocation analysis is the corpus linguistic go-to tool. Most simply defined, a collocation refers to two words co-occurring in close proximity to each other so frequently that this co-occurrence is above chance (McEnery, Hardie 2012: 122ff). Collocations contribute to meaning and can thus provide semantic insight into a word.[8] The results of the CQPweb's collocation tool are typically lists[9] of words that disproportionately often occur in the vicinity of a given query word. Again, with collocations a historic perspective is interesting for capturing changes or consistencies in meaning. To this end, the concept of 'c-collocates' (constant-collocates) was introduced by Gabrielatos and Baker (2008b) and is used to statistically filter historically consistent collocates of a word. The c-collocates for this study were calculated[10] with an independent online tool that uses data exported from CQPweb (available at: http://corpora.lancs.ac.uk/stats/toolbox.php)

The most important c-collocates calculated in this way are then subjected to close examination by studying concordance lines. In the CQPweb corpus tool the concordance lines display shows query items in a broader context, with, for example, up to a couple of sentences to the left and right of the item, as we can see in Figure 2 below. This shows part of the concordance line display in CQPweb for the collocates 'Frau' and 'Mädchen'. Concordance lines can thus be used to study the broader contexts of query terms.

Figure 2: Detail of the concordance lines display for the collocation of 'Frau' and 'Mädchen'

No.	Text	Solution 1 to 46 Page 1 / 1
1	jbav_1872_003_128	Stube trieb . Dort hatten die Gäste gewechselt , die festlich geputzten Frauen und Mädchen hatten das Feld geräumt und es den früher auf der
2	jbav_1881_012_196	Haus prüfend und segnend , die echte Beschützerin liebender Mädchen und emsiger Frauen . Entstellt ist sie wiederum inder Localsage von der Wildenburg . In
3	jbav_1886_017_405	vielleicht mit einziger Ausnahme des wirklich hübschen grünen Spitzhütchens der Mädchen und Frauen , zu verwischen drohte , durch einen Verein für Erhaltung der Volkstracht

Source: Author.

---

[8] For a detailed discussion of the metrics and tools used to calculate collocations see Brezina (2018).

[9] In more recent work graphical visualisations have been used in the form of collocation networks.

[10] For a discussion of what this statistic entails see the paper by McEnery, Brezina and Baker (2019).

That said, as this study is specific for the *AACJ* dataset the linguistic patterns uncovered by the statistical means used here may be unique to this study. However the method may 'trace language patterns and keywords in a way that contributes to the investigation of gender bias in sports discourses' (Ismail 2016: 171).

The concepts of nomination and representation in the DHA

In the DHA the concept of nomination refers to linguistic strategies of reference (self- and other) and is related to person reference or, as proposed by Hornscheidt (2006), appellation. Nomination entails linguistic strategies for indicating group membership to an in- or outgroup (like said person reference or personal pronouns, etc.). Constructivist pragmatics criticises the term 'person reference' because it presupposes that an entity that pre-exists outside language is linked to an also pre-existing word describing the entity. 'Appellation', on the other hand, evokes more than a simple reference and indicates that the usage of a certain word also functions as a linguistic appeal evoking such an identity. Every appellation thus also has an appeal function and evokes something in a person's mind and thus serves the concept of nomination in discourse. Different nominations in a text that, for example, point to a 'biological' person in everyday language produce this person with certain characteristics as a 'linguistic' person. In this sense, person appellations are mechanisms that establish the textual and intertextual identity of a person/group of people. Through repetition and citation, these fabrications become solidified and 'sedimented', creating an identity image of the person addressed/discussed in the text.

In a noun-centred language like German, it makes sense to study the concept of 'nomination' by first taking a look at nouns. Person appellation may include regular nouns and also names and name affixes ('Dr', 'Prof', etc.). Such an approach is by no means trivial to accomplish automatically since word formation processes as well as inflection and grammatical gender in personal appellation are complex and diverse. Particular attention must be paid here to how the problem of masculine generics is to be interpreted. Kotthoff and Nübling (2018) note the close connection between gender and case and also discuss the problematic of both in connection with category number as well. Grammatical gender assignment in German follows different principles, one of which is the semantic principle: the lexical meaning of the word attributes grammatical gender (sometimes also 'lexical gender'). Gender is also assigned by means of inflection, i.e. by adding a (feminine) suffix to a grammatically masculine word, e.g. *Arbeiter* ('worker') → *Arbeiterin* ('worker-FEM'). The set of words that use the semantic principle is more limited, whereas gender inflection is possible for many forms of personal appellation. This paper

focuses on the representation of women and studies two appellation forms that evoke female person(s).[11]

In order to grasp the representation of gender in nomination patterns there are some pre-requisites according to Kotthoff and Nübling (2018: 115): The person appellation as well as its degree of 'genderisation' or 'Genderisierungsgrad' (ibid.) are active simultaneously. This means that not every person reference noun is equally representative according to gender, as this is closely linked to stereotypes or biases concerning gendered roles.

Selected aspects of person appellations according to feminine gender and female persons are studied in more detail below. First, an overview of the metadata fromthe corpus is given in order to present a picture of gender representation in the composition of the corpus data. Then, with corpus-driven and corpus-based methods, the developments of some selected items are demonstrated by way of examples.

Findings
Gender representation in the composition of the corpus

Figure 3 shows the composition of the corpus based on the gender of the first authors of individual articles/texts.[12] Previous studies have found that the total percentage of news articles about female athletes is less than 10% of all such coverage (Caldwell 2016: 173). As can be seen in Figure 3, the number of female authors has increased in recent years. Altogether, 88.5% of all the articles in the *AACJ* were first authored by men and only 5.65% were first authored by women. As a source for the figure, the table of contents database from Alpenwort was used, which was derived from *corpus* texts during the corpus-building process and later corrected and enriched with further metadata. A closer look at the texts reveals that a large proportion of the texts written by female authors are from one special issue concerned with the topic 'women in the mountains'. Thus, the overall increase observed may not necessarily be a sign of an increasing overall representation of women authors in the journal, but rather signify the acknowledgement of feminist tendencies in society.

Figure 4 is slightly different in that it shows the number of articles in the corpus that directly mention women in the title versus the number of articles that were catalogued

---

[11] We assume that in the historical part of the corpus gender categories beyond the binary model will not be present at all in a medium like the *AACJ*.

[12] Only a relatively small number of articles (130 out of 3264 articles altogether) have multiple authors, and women are the second authors of 5 of these (3 different women). The authors of 5.85% of the articles are unknown or only an abbreviated form of the name is given, and the authors cannot be identified. This is predominantly a problem in the older issues and one can assume that most of the unidentifiable authors were probably men.

as *Frauenalpinimsus* ('women's alpinism'). This was achieved by comparison with another table of contents dataset provided by the Alpine Club's library that includes a list of library catalogue keywords.

Figure 3: Number of articles written by women, men, unknown authors
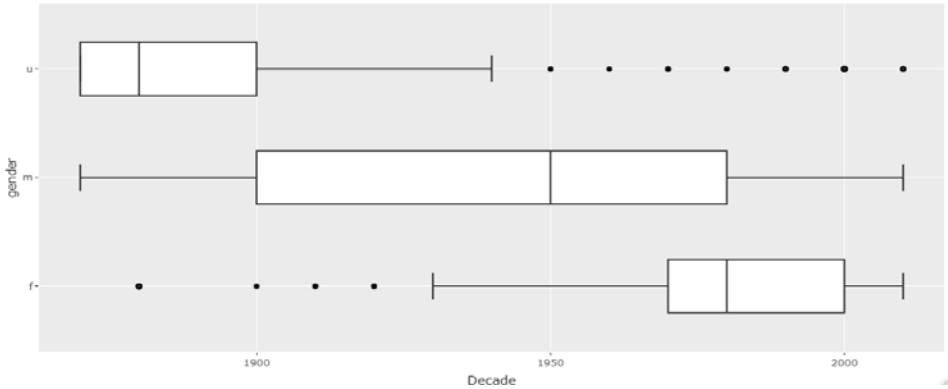
Source: Author.

Figure 4: Articles on women (according to the title) vs articles with the keyword Frauenalpinismus ('women's alpinism')
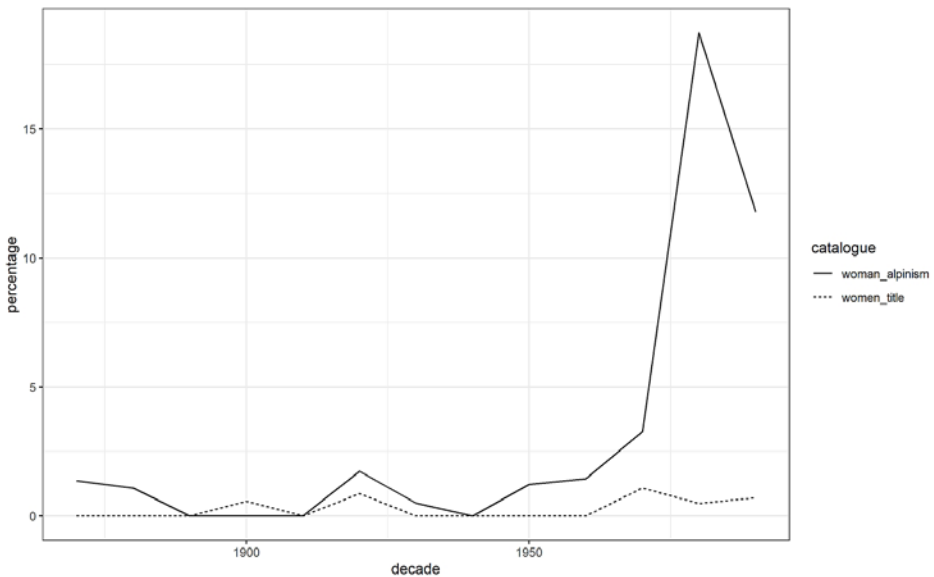
Source: Author.

Gender representation in the corpus texts
Overview of key items

A key item analysis was the first step in the investigation of gendered nomination in the corpus texts. A keyness and down sampling approach, as described in Posch (2022; Rampl, Posch 2022), was applied using trend mapping, where each decade was used as

1. a corpus of interest and then
2. compared to a much larger reference corpus that consisted of all the other decades by using CQPweb's keywording tool.

The following adjustments were made to the key item calculation method of the CQPweb version of the corpus:

CQPweb-settings: Log Ratio (with 0.1% significance filter, adjusted LL threshold = 33.03); items must have minimum frequency 3 in list #1 and 3 in list #2.

This resulted in one list per decade with more than 14,000 keywords. Feminine person appellations were extracted automatically from the lists with a manual cross-check.

Figure 5: Feminine person appellation key items over time[13]
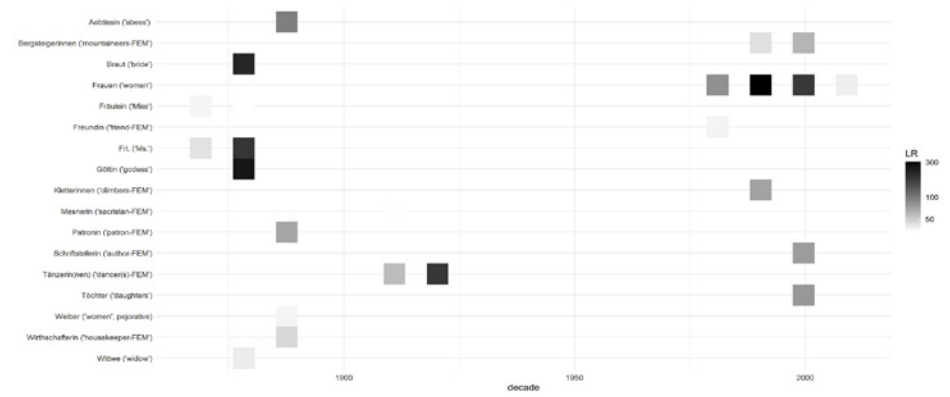
Source: Author.

Figure 5 shows all the person appellation items that were key at some point in the corpus. Only five of the words[14] are salient key items in more than one decade: *Bergsteigerinnen* ('mountaineers-FEM'), *Frau(en)* ('woman/women'), *Fräulein* ('Miss'),

---

[13]   Aebtissin, Mesnerin, Wirthschafterin, and Wittwe are the historical spellings of these words.
[14]   This includes the consolidated word class 'person name' with individual names; a typically female person name is key in almost every decade.

and *Tänzerin(nen)* ('dancer(s)-FEM'). *Fräulein* is historically a form of address that from the 19th century on was used for unmarried women ('Frau', in Wolfgang Pfeifer et al., Etymologisches Wörterbuch des Deutschen (1993), <https://www.dwds.de/wb/etymwb/Frau>, abgerufen am 20.03.2022.). *Fräulein* is a key item in two consecutive periods in the early decades before 1900. It is the diminutive form of 'Frau', the regular word for woman. The word *Fräulein* is used only 356 times overall (fpmw 17.89) in 125 different texts mostly until the 1930s. It is also used 26 times (fpwm 18.48) in the 2000s decade. Here, in contrast, the word *Fräulein* is not used in this obsolete sense, but is part of a toponym, such as in *Wildes Fräulein*, or refers to the name of mythical creatures associated with high mountains – for example, the *Salige Fräulein* in articles about such myths.

Figure 5 further shows a difference in the types of appellations that occur over time. Some of the feminine items used in the first half of the corpus are words with lexical genders (= gender assignment by semantics and a high degree of genderisation) such as *Weiber* ('women'), which means 'adult woman' as well as 'wife'. Two further words are roughly connected to marital status as well: *Wittwe* ('widow') [historical spelling] and *Braut* ('bride'). Four words refer to religious beliefs and/or a clerical function: *Göttin* ('goddess'), Patronin ('patroness'), *Aebtissin* ('abbess') [historical spelling], *Mesnerin* ('sacristan-FEM') [historical spelling]. The word *Jungfrau* ('virgin, maiden') is a reference to the Virgin Mary, but it much more frequently occurs in reference to the name of the famous mountain in Switzerland.

The only person appellation in the first half of the corpus that is formed not on the basis of lexical gender but via gender inflection is *Tänzerin* ('dancer-FEM'), which is key in two consecutive decades before the 1950s. It is apparent that none of the keywords referring to a female person before the 1950s has an obvious semantic connection to the activity of mountaineering itself and they display high degrees of genderisation.

The picture changes in more recent years. The word *Frau(en)* ('woman/women') is a keyword in three consecutive periods until the 2000s, which clearly indicates the increasing importance of the word. In both the 1990s and the 2000s the word *Bergsteigerinnen* ('mountaineers-FEM') is a keyword, while the word *Kletterinnen* ('climbers-FEM') is key in one decade, and both these appellation forms have a clear connection to the activity of mountaineering. Two other inflected feminine forms are also key in one decade each: *Schriftstellerin* ('writer-FEM'), *Freundin* ('friend-FEM') (i.e. 'girlfriend'). Last but not least, *Töchter* ('daughters') is the only word with a lexical gender that is also key in one of the later decades.

The development of other feminine forms: a selection

A first glimpse at the representation of women in the Austrian Alpine Club Journal was obtained by looking at how, when, and in what context the above keywords are used to refer to/evoke women. Second, some other quasi-synonymous forms were queried to get a broader picture. To research this a search for lemmata was conducted in the CQPweb tool of the Alpenwort corpus. The query terms included the key items and were extended for other items with lexical gender:

[lemma = "Frau|Weib|Dame|Fräulein|Mädchen|Magd|Mama|Mutter|Oma|Tante| Tochter"]

The query resulted in 7,147 matches in 1,493 different texts, which represents a normalised frequency of 359.14 instances per million words. For scale, the search for the equivalent (if available) or similar semantically masculine words resulted in 16,833 matches in 2,269 texts, which amounts to a frequency of 845.86 instances pmw[15].

Figure 6: Lexical gender items diachronically
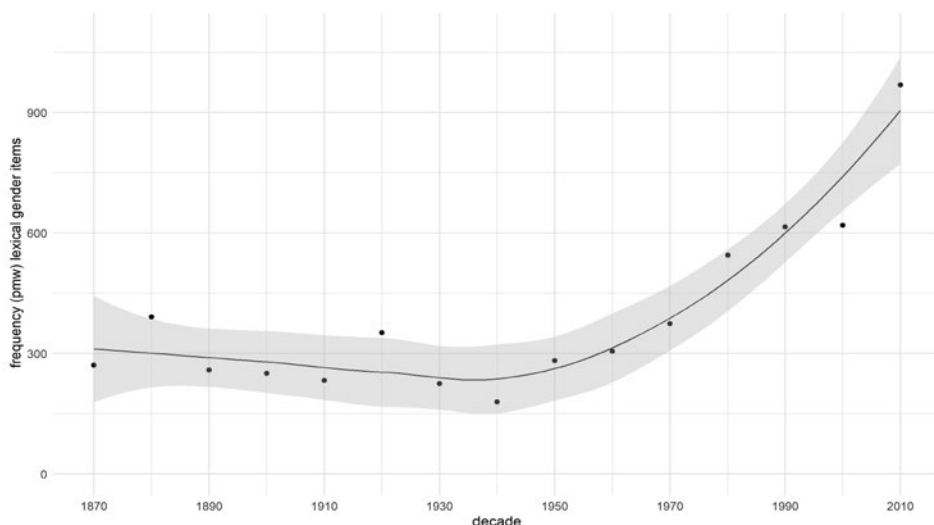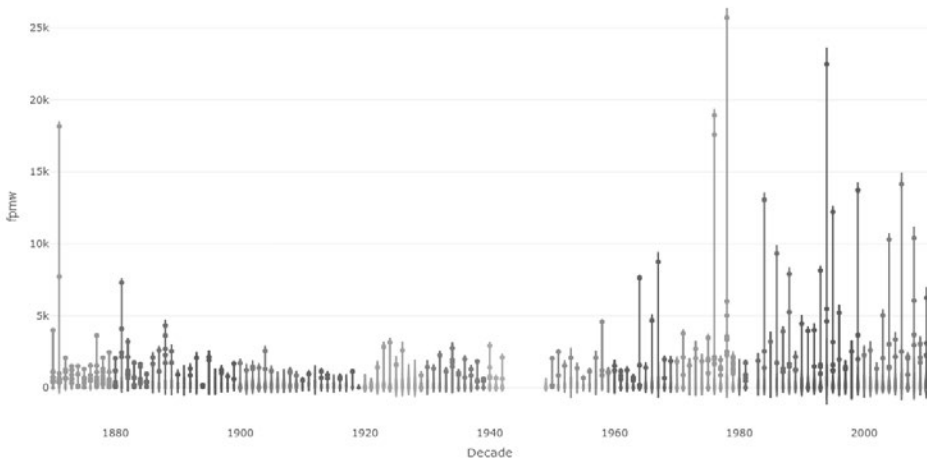

Source: Author.

Figure 6 shows an almost level line from 1870 to the decade of the 1950s in the usage of lexically feminine items (fmpw). There is a slightly decreasing trend with two outliers in the decades 1880 and 1920, respectively. Then a steep increase begins and continues until the decade of the 2000s.

---

[15] [lemma="Mann|Herr|Bub|Knecht|Bursch|Papa|Vater|Opa|Onkel|Sohn"]

Figure 7: Fpmw per individual year (=book)

A closer look at the individual texts reveals that there are some extreme outliers and that the number of these outliers increases in particular from the late 1970s on (Figure 7). The text with the most extreme fpmw of the queried words is from 1978 (jbav_1978_103_246)[16] and it is called *Frauen in den Bergen* ('women in the mountains').

The only early text in which we find a rather exceptional fpmw-value is from 1871 (jbav_1871_002_797) and is titled *Das Schrötterhorn* (the name of a mountain). It is a very short text (55 words) in which the word *Frau* is used once – see example (1).

1. Die erste Ersteigung dieses Gipfels erfolgte im August 1870 durch Herrn und **Frau** Franz Gröger aus Wien, unter Führung von Joh. und Jos. Pinggera. [*The first ascent of this peak was made in August 1870 by Mr and **Mrs** Franz Gröger from Vienna, guided by Joh. and Jos. Pinggera.*]

The example uses the word *Frau* in its meaning as a form of address. The woman described is not even mentioned by name but is subsumed under the name of her husband, *Franz Gröger.* As an interesting sidenote, the website for first ascents by the *Alpenverein* (http://www.alpinwiki.at/portal/navigation/routen/routedetail.php?route=29384) does not even co-list the unknown wife of Franz Gröger as first ascender of the *Schrötterhorn*, and only Franz Gröger and the two guides are named.

---

[16] There is an amusing coincidence here, as the author of this text is a Claudia Posch, whom I do not know.

2. Schalfkogl: 2 Partien und 6 Touristen, darunter 2 **Damen** (**Frau** v. Statzer aus Wien und als erste Ersteigerin des Schalfkogls **Frau Dr.** Scholle aus Berlin) [*Schalfkogl: 2 parties and 6 tourists, including 2 **ladies** (**Mrs** v. Statzer from Vienna and, as the first **female climber** of the Schalfkogl, **Mrs** Dr Scholle from Berlin)]*

3. Weisskugel: 4 Partien und 6 Touristen, darunter **Frau** v. Statzer aus Wien, die erste **Dame** auf der Weisskugel [*Weisskugel: 4 parties and 6 tourists, including **Mrs** v. Statzer from Vienna, the first **woman** on the Weisskugel]*

4. Ebenso ausserordentlich war die zweite Ersteigung der Finailspitze durch den genannten Herrn mit seiner Mutter, **Frau** von Statzer. [*Equally extraordinary was the second ascent of Finailspitze by the aforementioned gentleman with his mother, **Mrs** von Statzer.]*

Another early outlier is the text jbav_1871_002_713 titled *Venter Touristenkalender, 1870* (Vent [a placename] touristic calendar, 1870), which has a total of 7 hits in a text with 905 words. The examples (2–4) highlight the words used: *Damen* ('ladies'), *Frau* (in the sense of 'Mrs'), and *Mutter* ('mother'). Again, *Frau* is used as form of address, and in example (2) two women are even named as ascenders.[17] The women are highlighted as extraordinary by using the word *darunter* ('among'). The general person appellation used in this example is *Damen*, which was the unmarked form for women at the time. The name *Frau v. Statzer* (lit. 'Mrs from Statzer') is mentioned again in example (3), again with the same form of emphasis (*darunter*). It is also mentioned that this woman was the first female ascender of the mountain *Weißkugel*. In example (4) the same woman is mentioned again in the form of a comitative attribute of a man (*des genannten Herrn* 'the aforementioned gentleman'), whose mother she is. The women mentioned in those lists of alpinist achievements are not mentioned anywhere else in the corpus and their identities cannot be further traced in the data.

Table 1 shows the other texts with an fpmw of more than 5k. Four of the texts with the most extreme usages of a semantic gender form are by the author Ingrid Runggaldier, a publicist who specialised in writing about female achievements in mountaineering. Two texts are from one author, Liselotte Buchenauer. The other texts are written by different authors. Fifteen of the texts have an explicit focus on women in mountaineering and discuss the topic of women in this typically male domain. It is thus no surprise that those texts use more feminine person appellations as this reflects the overall increase in the number of texts concerned with women in alpinism.

---

[17] 'Sclialfkogls' is an error produced by text recognition. This happens in particular with toponyms.

Table 1: Articles with outliers of fpmw above 5

Text ID (jbav_)	Title	Author	wit[1]	hits	fpmw
1978_103_246	*Frauen in den Bergen*	Claudia Posch	1011	26	25,717.11
1994_118_59	*Frauenalpinistik*	Dagmar Wabnig	2,936	66	22,479.56
1976_101_181	*Frauen am Berg*	Lia Hörmann	634	12	18,927.44
1871_002_797	*Das Schrötterhorn*	Franz Gröger	55	1	18,181.82
1976_101_48	*Frauen im Alpenverein*	Richard Grumm	1,534	27	17,601.04
2006_130_131	*Die Frau im Fels Die Anfänge des Frauenbergsteigens in Südtirol*	Ingrid Runggaldier	5,867	83	14,146.92
1999_123_103	*Ist der Berg männlich? Frauen und Bergsport*	Dagmar Nedbal	4,806	66	13,732.83
1984_108_54	*»Was, das Madl will führen…?«*	Judith Huber	6,042	79	13,075.14
1995_119_161	*Frauenbergsteigen. Auf der Suche nach einer vergessenen Seite der alpinen Geschichte.*	Anette Köhler	5,642	69	12,229.71
2008_132_73	*Mein Gott! Wie war das alles schön und wild und schrecklich! Die Berge in den Augen der französischen Schriftstellerin George Sand (1804-1876)*	Ingrid Runggaldier Moroder	4,132	43	10,406.58
2004_128_27	*Ötzis Göttinnen Auf den Spuren von Sagen zu Stätten matriarchaler Kulturen in den Ötztaler Alpen*	Hans Haid	6,199	64	10,324.25
1986_110_100	*Emanzipation alpin*	Joëlle Kirch	5,253	49	9,328.00
1967_092_173	*Vierhundert Jahre Bergsteigerinnen*	Felicitas von Reznicek	5,027	44	8,752.74
1993_117_95	*Sie gehen, wie es ihnen taugt*	Lydia Kraus	4,048	33	8,152.17
1988_112_83	*Berge - Bücher - Begegnungen*	Liselotte Buchenauer	5,311	42	7,908.12
1871_002_713	*Venter Touristenkalender, 1870*	Franz Senn	905	7	7,734.81
1964_089_156	*Drei große Bergsteigerinnen. Mira Marko Debelakova-Derzaj, Laibach (Ljubljana) (1904—1948), Grete Rieder-Großmann, Graz (1913—1948), Maria Kampitsch, Leoben (191 — 1956)*	Liselotte Buchenauer	5,094	39	7,656.07
1881_012_196	*Die Göttin Bercht-Holda und ihr Gefolge. Beiträge zur Erklärung ihres Cultus und der darauf bezüglichen Volksbräuche.*	Dr. Ludwig Freytag	18,450	135	7,317.07

2010_134_107	„Meinem Mann zuliebe“ … die Höhenbergsteigerin Hettie Dyhrenfurth	Ingrid Runggaldier	6,874	43	6,255.46
2008_132_104	Von wegen Eisprinzessin Erfolgreich in einer Männerdomäne: Ines Papert	Karin Steinbach Tamutzer	7,583	46	6,066.20
1978_103_155	Erste Winterbegehung der Matterhorn-Nordwand 1978 in Damenseilschaft	Wanda Rutkiewicz	1,165	7	6,008.58
1994_118_131	al hamdu Allah	Anna Lauwaert	727	4	5502.06
1988_112_153	Als Frau in der Todeszone	Henriette Eberwein	6,262	33	5269.88
1996_120_177	Aubet, Cubet, Quere	Henriette Klier-Prochaska	7,095	37	5214.94
2003_127_134	Felskletterer - mit Licht gezeichnet Rudolf Köbach und die Entwicklung der Bergfotografie im Elbsandsteingebirge	Helmut Schulze	2,976	15	5040.32
1978_103_218	Der Mensch, ein Wanderer	Raul O. Mendoza	597	3	5025.13

Source: Author.

The most frequent semantically gendered lemma in the corpus by far is *Frau* ('woman'), which, as noted above, can be used as both a form of address and in its appellative sense as a word for woman. The table below presents the frequency breakdown of search items from CQPweb, which shows the number of occurrences of search items in the corpus as a whole. The word *Frau* in both singular and plural accounts for 52.03 percent of all hits in the search. The next most frequent word is *Mutter* at 10.29 percent, followed by *Mädchen* ('girl') at 8.71 percent. All the other terms each are under 5%.

Table 2: Frequency breakdown of lexical gender words

No.	Search result	Translation	No. of occurrences	Percent
1	Frau	woman	2629	36.78
2	Frauen	women	1319	18.46
3	Mutter	mother	781	10.93
4	Mädchen	girl	661	9.25
5	Tochter	daughter	372	5.2
6	Damen	ladies	268	3.75

7	Fräulein	miss	258	3.61
8	Weib	woman (pejorative)	179	2.5
9	Dame	lady	175	2.45
10	Weiber	women (pejorative)	171	2.39
11	Frl.	ms.	84	1.18
12	Magd	maid	48	0.67
13	Mägde	maids	37	0.52
14	Töchtern	daughters [dative]	33	0.46
15	Müttern	mothers [dative]	28	0.39
16	Tante	aunt	26	0.36
17	Mädchens	girl [genitive]	25	0.35
18	Mama	mum	21	0.29
19	Fräuleins	miss [genitive]	14	0.2
20	Mägden	maids [dative]	8	0.11
21	Oma	granny	7	0.1
22	Mamas	mums	1	0.01
23	Omas	grannies	1	0.01
24	Tanten	aunts	1	0.01

Source: Author.

Another directed search was conducted for person appellation using gender inflection. Here only words that semantically refer explicitly to mountaineering were queried:

[lemma=".*ergsteigerinl.*letterinl.*lpinistinl.*raxlerin"][18]

This query produces simplexes as well as compounds and resulted in a list of 359 hits in 140 texts with a normalised frequency of 18.04 instances per million words. Again, for scale the equivalent masculine appellation terms resulted in much higher numbers: 12,987 matches in 2,200 different texts (i.e. 652.60 fpmw). The search found 19 different types of person appellations referring to female alpinists,[19] whereas the search for the masculine forms yielded 75 types.

---

[18] This searches for simplexes and compounds with feminine forms for: 'mountaineer', 'climber', 'alpinist', 'mountain climber'.

[19] Bergsteigerinnen, Bergsteigerin, Kletterin, Kletterinnen, Alpinistin, Alpinistinnen, Höhenbergsteigerin, Sportkletterin, Wettkampfkletterin, Expeditionsbergsteigerin, Extrembergsteigerin, Profibergsteigerin, Eiskletterin, Eiskletterinnen, Himalajabergsteigerin, Leistungsbergsteigerin, Skialpinistinnen, Spitzenkletterin, Spitzenkletterinnen.

Figure 8: Peaks and troughs analysis of the use of [lemma=".*ergsteigerinl.*lette-rinl.*lpinistinl.*raxlerin"]
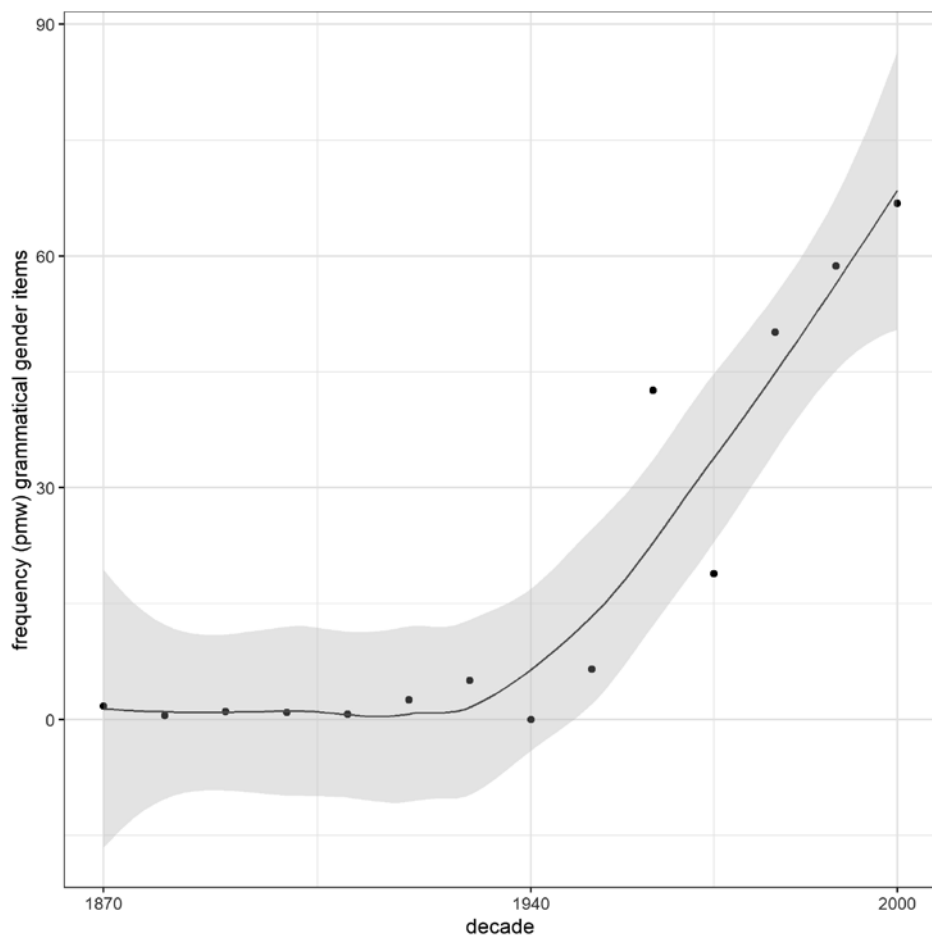
Source: Author.

The peaks and troughs technique (Figure 8) shows a departure from zero and a steady increase of uses with a peak in the 1970s and slight flattening of the curve towards the end of the scale. The highest fpmw is 5,434.78 in the text *Sie gehen, wie es ihnen taugt* ['They will climb how it suits them'] from 1993). None of the 24 texts with frequencies above 1,000 fpmw for the queried items is from earlier than 1964.

Table 3: Texts with words inflected for feminine gender

Text ID (jbav_)	Title	Author	wit	hits	fpmw
1993_117_95	*Sie gehen, wie es ihnen taugt*	Lydia Kraus	4,048	22	5,434.78
1964_089_156	*Drei große Bergsteigerinnen.*	Liselotte Buchenauer	5,094	26	5,104.04
2006_130_131	*Die Frau im Fels Die Anfänge des Frauenbergsteigens in Südtirol*	Ingrid Runggaldier	5,867	26	4,431.57
1978_103_155	*Erste Winterbegehung der Matterhorn-Nordwand 1978 in Damenseilschaft*	Wanda Rutkiewicz	1,165	4	3,433.48
1978_103_246	*Frauen in den Bergen*	Claudia Posch	1,011	3	2,967.36
1967_092_173	*Vierhundert Jahre Bergsteigerinnen*	Felicitas von Reznicek	5,027	13	2,586.04
1986_110_100	*Emanzipation alpin*	Joëlle Kirch	5,253	12	2,284.41
2007_131_5	*Vorwort*	Walter Theil	959	2	2,085.51
2007_131_51	*Elbrus-Besteigung*	Matthias Pinn	4,132	8	1,936.11
2007_131_53	*»Chaddar« - Straße aus Eis*	Andreas Mühlleitner	4,262	7	1,642.42
2007_131_57	*Klettern zwischen Reisfeldern, Affen und heiligen Kühen Bericht über die Kletterfahrt der AVS-Jugend nach Südindien*	Bernhard Mock	634	1	1,577.29
2008_132_73	*Von wegen Eisprinzessin Erfolgreich in einer Männerdomäne: Ines Papert*	Karin Steinbach Tamutzer	3,223	5	1,551.35
2009_133_109	*May Norman-Neruda (1867-1945)*	Ingrid Runggaldier	6,042	9	1,489.57
1976_101_181	*Frauen am Berg*	Lia Hörmann	5,708	8	1,401.54
1994_118_181	*Alpinismus international*	Dieter Elsner	3,141	4	1,273.48
1984_108_54	*»Was, das Madl will führen...?«*	Judith Huber	1,710	2	1,169.59
1989_113_119	*Wie die Bergsteiger reden*	Joëlle Kirch	5,155	6	1,163.92
1989_113_57	*Der große Swing*	Elmar Landes	6,149	7	1,138.40
1972_097_83	*Der weiße Elefant und die Bergsteiger*	Günter Hauser	922	1	1,084.60
2008_132_59	*Gaby Hupfauer Berge - eine Leidenschaft fürs ganze Leben*	Gaby Funk	3,699	4	1,081.37
1990_114_246	*Der konsumierte Berg*	Richard Goedeke	6,558	7	1,067.40

2005_129_13	»Bestürzend und beeindru*ckend!*« *- die Ausstellung »Gletscher im Treibhaus« im Alpinen Museum des DAV*	Friederike Kaiser	4,806	5	1,040.37
2005_129_131	*Mehr Seil, bitte! Fingierte Tagebucheinträge einer Einsteigerin*	Dennis Cramer	2,936	3	1,021.80
2005_129_133	*Keine richtigen Berge! Die vielgestaltige Voralpenlandschaft um Tegernsee und Schliersee*	Stefan König	6,874	7	1,018.33
2006_130_151	*Auf der Suche nach dem Kick Wie risikobereit sind jugendliche Alpinsportler?*	Wolfgang Wahl	4,048	22	5,434.78
1991_115_55	*Bergsteigen oder nicht?*	Elmar Landes	5,094	26	5,104.04
1994_118_59	*Frauenalpinistik*	Dagmar Wabnig	5,867	26	4,431.57
2010_134_107	»Meinem Mann zuliebe« ... die Höhen*bergsteigerin Hettie Dyhrenfurth*	Ingrid Runggaldier	1,165	4	3,433.48

Source: Author.

Words in context – collocations

The collocations for the lemma *Frau* were calculated with CQPweb using the following settings:

Lemma: /Frau/ **Statistic:** Z-Score[20], **collocation window:** 3L to 3R, Freq (node, collocate): min. 5; Freq (collocate) min: 5;

In many cases *Frau* collocates with a name, usually the individual person names of women, sometimes the collocation is also with the name of a man in the context of -/+ 3 words before or after the node – e.g. *Hans Pfann und Frau Eleonore Noll--Hasenclever.* The most relevant collocation is with the name *Hitt*, which is a mountain name – *Frau Hitt* (Lady Hitt) – and thus cannot be counted as a representational form for a person reference.[21] The second collocation is with the name-part Noll,

---

[20] The Z-score is a measure whose results reflect a combination of significance (amount of evidence) and effect size (strength of connection), producing a compromise ranking relative to MI (effect size) and LL (significance).

[21] The mountain name *Frau Hitt*, a mountain on the doorstep of Innsbruck, refers to a mythical figure, a petrified giant queen, who is associated with negative character traits stereotypically associated with women.

and it is used in reference to the very active Alpinist Eleonore Noll-Hasenclever, who gained some recognition in the 1920s. The third collocation is with the possessive *meine* ('my'); other word forms / possessives are also among the top 50 collocations, such as *meiner* ('my' – dative, genitive); *seiner* ('his' – dative, genitive). On rank five the collocation is with the plural form *Männer* ('men'), as, for example, in the binominal *Männer und Frauen* or, less frequently, *Frauen und Männer*. Ranks 6 to 21 are again mostly names, and rank 22 is the word *Kinder* ('children').
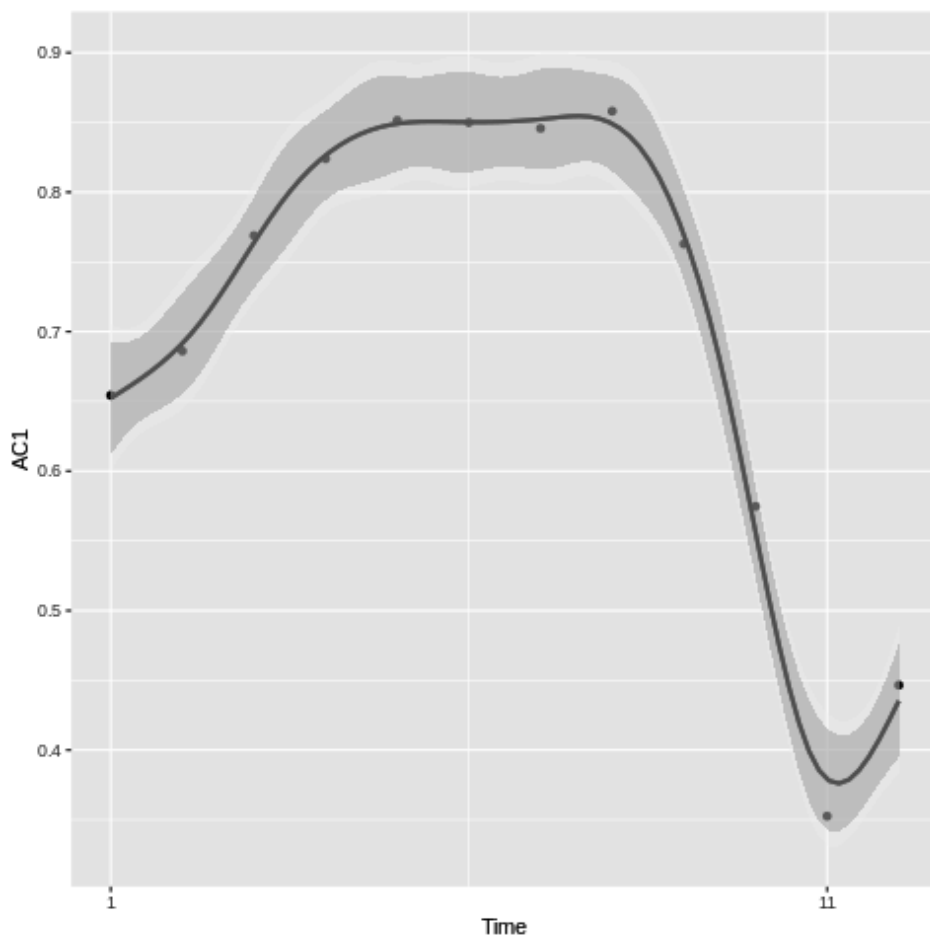
C-collocates:

To retrieve the constant collocations of a query term in a corpus, McEnery, Brezina and Baker (2019) suggest Usage Fluctuation Analysis (UFA). They distinguish between collocates that are relatively consistent across an overall period of time (consistent collocates), those that are consistent during a certain period but then fall out of use (terminating collocates), those that became consistent during the overall period (initiating), and those that 'attach themselves only briefly to a word' in the period (transient collocates) (McEnery et al. 2019: 417).

For this paper the query term *Frau* was investigated using the UFA-tool available at Lancaster Stats Tools online (http://corpora.lancs.ac.uk/stats/toolbox. php?panel=6&tab=3). The output of the tool is a graph 'showing convergence, or divergence, between collocates within sliding windows moving through time'.

The UFA graph displays time slots on the *x* axis and the calculated agreement statistics on the *y* axis (Gwet's AC1). The points for most of the graph are rather closer to 1 (AC1 scale 0–1), which means that the collocates do not really change between the consecutive points compared. There is a gradual shift in the early periods towards a more stable plateau. There is a sudden dip in the graph towards the later periods (11 i.e. the decade of the 1970s). The lower points mean that at this point in time there is rather less agreement between the consecutive points than in other pairwise comparisons in the analysis. This coincides with the increasing inclusion of female authors writing about women in mountaineering in the journal and signifies a change in the collocations of the query term. Peaks indicate similarity in the sets of collocates. On the *y* axis 0 indicates agreement by chance, whereas 1 indicates perfect agreement. Thus, a dip means that the collocations are not as strong in the later period.

Figure 9: UFA analysis of the lemma Frau ('woman')

Source: Author.

A closer look at the collocation output of the tools will enable a qualitative interpretation of these findings.

SUMMARY: The usage fluctuation analysis (UFA) identified 171 unique collocates across 14 periods. Full table with collocates is also available.

CONSISTENT: als (2-14), das (2-14), den (2-14), der (1-14), die (1-14), eine (2-14), in (1-14), **meine** (2-14), **meiner** (3-14), mit (1-14), **seine** (2-14), und (1-14), von (1-14),

INITIATING:

TERMINATING:

TRANSIENT: ich, 4-14, 64.3%; im, 1-14, 64.3%; auch, 2-14, 64.3%; sich, 2-14, 64.3%; bei, 2-14, 64.3%; dem, 2-14, 57.1%; zu, 2-14, 57.1%; seiner, 2-14, 57.1%; auf, 4-14, 57.1%; einer, 2-14, 50%; hatte, 4-14, 50%; war, 5-14, 50%; aus, 1-14, 50%; an, 3-14, 42.9%; es, 2-14, 42.9%; nach, 2-14, 42.9%; noch, 3-14, 35.7%; wie, 4-14, 35.7%; ein, 4-14, 35.7%; ist, 2-14, 35.7%; er, 1-14, 35.7%; erste, 7-14, 35.7%; so, 2-14, 35.7%; nicht, 2-14, 35.7%; sie, 2-14, 35.7%; für, 7-14, 35.7%; mädchen, 2-13, 35.7%; männer, 4-14, 35.7%; hat, 9-14, 28.6%; am, 11-14, 28.6%; kinder, 11-14, 28.6%; immer, 11-14, 28.6%; man, 10-14, 28.6%; durch, 3-14, 28.6%; ihre, 3-13, 28.6%; einem, 2-14, 28.6%; junge, 10-14, 28.6%; nur, 12-14, 21.4%; keine, 12-14, 21.4%; aber, 2-13, 21.4%; bergsteigen, 12-14, 21.4%; männern, 12-14, 21.4%; zwei, 12-14, 21.4%; unser, 1-3, 21.4%; um, 11-14, 21.4%; dass, 2-14, 21.4%; waren, 7-14, 21.4%; ihren, 7-14, 21.4%; sind, 12-14, 21.4%; des, 10-14, 21.4%; mann, 12-14, 21.4%; gipfel, 12-14, 21.4%; oder, 2-13, 21.4%; lieben, 1-4, 21.4%; ob, 12-13, 14.3%; schon, 13-14, 14.3%; sehr, 12-14, 14.3%; uns, 6-13, 14.3%; unter, 13-14, 14.3%; anna, 1-2, 14.3%; v., 1-2, 14.3%; wenn, 12-13, 14.3%; beiden, 13-14, 14.3%; berg, 12-13, 14.3%; wird, 13-14, 14.3%; bis, 13-14, 14.3%; wurde, 13-14, 14.3%; während, 3-4, 14.3%; zum, 12-13, 14.3%; diese, 13-14, 14.3%; dr., 4-11, 14.3%; drei, 13-14, 14.3%; einen, 12-13, 14.3%; gibt, 12-13, 14.3%; haben, 13-14, 14.3%; herr, 3-4, 14.3%; herrn, 2-3, 14.3%; k., 1-2, 14.3%; kaufmann, 1-2, 14.3%; heute, 12-13, 14.3%; klettern, 12-14, 14.3%; marie, 1-2, 14.3%; mehr, 12-13, 14.3%; kind, 13-14, 14.3%;

The numbers after the collocation item indicate the period in which a particular item occurs as a collocate and the percentages show the proportion of periods in which the collocate occurred out of all the periods considered. Only collocates with a double-digit percentage were included in the table and the collocates were sorted according to percentages.

There are three interesting meaningful collocates that qualify as *consistent* – namely, the possessives *meine* ('my' – nominative, accusative), *meiner* ('my' – genitive, dative), and *seine* ('his' – nominative, accusative). The collocation with two forms of the possessive 'my' is exclusively connected to the specific meaning aspect of the word *Frau* as 'wife' and places the usage of the lemma 'Frau' in terms of the dependency of a husband, as the examples in Table 4 indicate.

GENDER **A** VÝZKUM

Table 4: Concordance lines of meine Frau, thinned random selection, non-reproducible

Text	Concordance line
jbav_1909_040_62	nervenzerrüttender Fahrt kamen **meine Frau** und ich Ende März 1905 nach (*nerve-wracking drive, my wife and I arrived in late March 1905 in*)
jbav_1914_045_9n	Kaltenbach, **meine Frau** und ich nebst Mirsa-baschi, Mirsa Achram (*Kaltenbach, my wife and I together with Mirsa-baschi, Mirsa Achram*)
jbav_1925_056_205	Da, ein Schrei, und kopfüber gleitet **meine Frau** den Eishang hinab (*There, a scream, and headlong my wife glides down the ice slope*)
jbav_1926_057_107	begleitete uns und brachte **meine Frau**, welche der Ruhe bedürftig war (*accompanied us and brought my wife, who was in need of rest*)
jbav_1927_058_118	Mit seiner überredenden Begeisterung infizierte er **meine Frau** und mich (*With his persuasive enthusiasm he infected my wife and me*)
jbav_1967_092_198	Rainer und die Hochträger waren ein Stück vorausgegangen, **meine Frau** und ich (*Rainer and the high altitude porters had gone a bit ahead, my wife and I*)
jbav_1990_114_38	Wir, **meine Frau** und ich, hatten eben unseren Gebietsführer (*We, my wife and I, just had our area guide*)
jbav_1999_123_48	Mit Tränen in den Augen saß **meine Frau** wenige Wochen später vor dem Fernseher (*With tears in her eyes, my wife sat in front of the television a few weeks later*)
jbav_2010_134_68	Im Juni 2007 flogen **meine Frau** und ich mit einer kleinen Gruppe gleichgesinnter (*In June 2007 my wife and I flew with a small like-minded group*)

Source: Author.

Figure 10: Distribution of 'meine Frau' in the corpus (fpmw per decade)
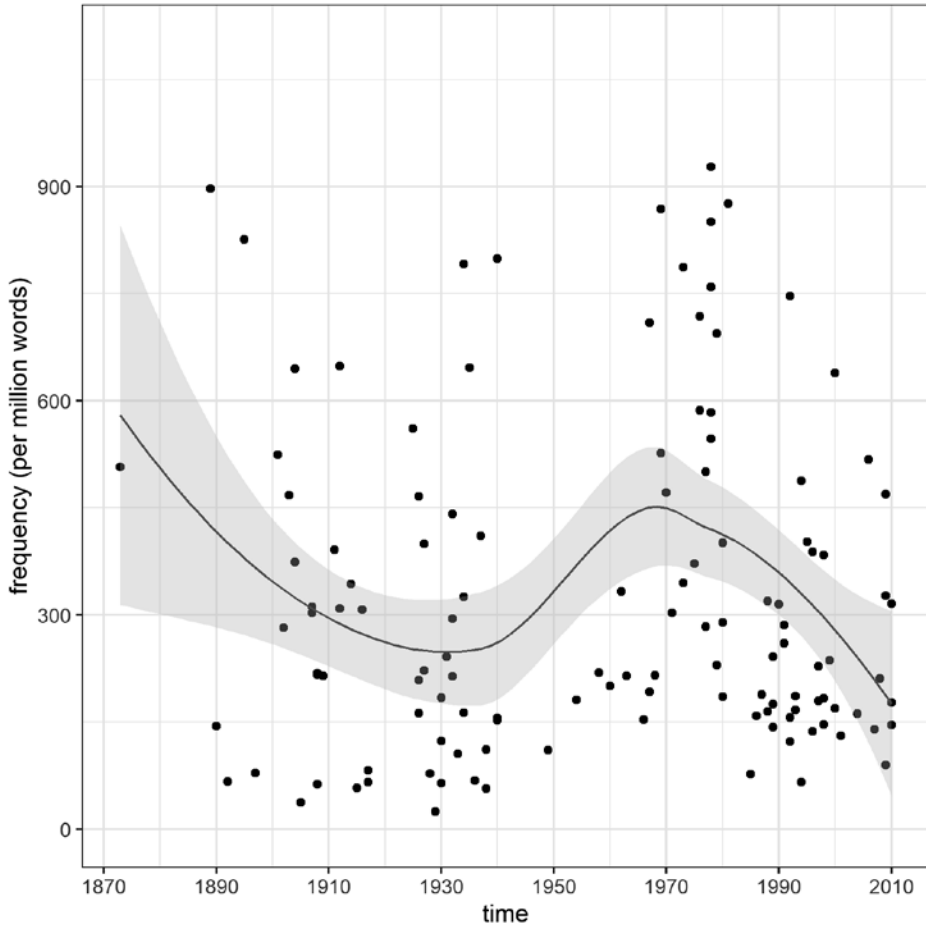
Source: Author.

Figure 10 shows that the combination *meine Frau* ('my wife') sets out from a higher fpmw and then moved into a flat trough, which is followed by an increase and peak around the 1970s and a sharper drop and decrease after that.

There are no initiating or terminating collocates of the lemma *Frau* in the overall corpus. A large number of collocates were categorised as *transient*, which typically reflects certain topics discussed or discourse fluctuations (McEnery et al. 2019: 420). Some of the more frequent collocations shall be briefly discussed below:

The query lemma *Frau* collocates with *Mädchen* ('girl/girls') in 35.7% of the decades investigated (decades 2–3). For example, the concordance lines show that this

collocation is usually a binominal of *Mädchen und Frauen* or *Frauen und Mädchen* in the 1880s (period 2) as well as in the 1990s (period 13), as the selected examples below (Table 5) show:

Table 5: Collocation of Frau with Mädchen

Nr	Text	Concordance line
1	jbav_1881_012_196	und segnend, die echte Beschützerin liebender **Mädchen** und emsiger **Frauen** (*and blessing, the real protector of loving **girls** and busy **women***)
5	jbav_1888_019_252	die man bei einer grossen Anzahl von **Mädchen** und **Frauen** findet, woraus gewiss auch ein Schluss auf die Unverdorbenheit (*which one finds in a large number of **girls** and **women**, from which one can certainly draw a conclusion about the purity*)
8	jbav_1999_123_103	Schon ermutigen einige Firmen **Mädchen** und junge **Frauen**, in vermeintliche Männerdomänen aufzubrechen (*Already, some companies are encouraging **girls** and young **women** to break into supposedly male domains*)
7	jbav_1999_123_10	Schar von 40 mutigen Männern, angefeuert von den zurückbleibenden **Frauen** und holden **Mädchen** (*A crowd of 40 brave men, cheered on by the **women** and fair **maidens** who stayed behind*)

Source: Author.

On the same level, *Frau* also collocates with *Männer* ('men', 35.7%, decades 4–14). Again, this mostly appears in the context of binominals like *Männer und Frauen* ('men and women'). As the examples below (Table 6) show, in earlier texts the binominal appears not to occur in the context of mountaineering at all but appears rather in cultural or ethnic descriptions as in (1). In more recent volumes there is a shift towards concern with differences and similarities of men and women in climbing, as those are, as mentioned before, the texts that are specifically concerned with women in mountaineering (2, 3).

Table 6: Collocation of Frau with Mann/Männer

Nr	Text	Concordance line
1	jbav_1909_040_196	am Ufer und am Strande ist das ganze Dörflein, **Männer** und **Frauen**, jung und alt beschäftigt, mächtige Reiserhaufen zu Bündeln zu ordnen (*On the shore and on the beach the whole village, **men** and **women**, young and old, are busy arranging mighty piles of sticks into bundles*)
2	jbav_2003_127_97	schätzt, dass **Männer** und **Frauen** über 70 Grad 9/9+ klettern könnten (*estimates that **men** and **women** could climb over 70 degrees 9/9+*)
3	jbav_2008_132_59	setzen sich dabei auch nicht so unter Leistungsdruck wie die **Männer**. **Frauen** muss es Spaß machen (*They also don't put themselves under as much pressure to perform as the **men**. It has to be fun for **women***)

Source: Author.

Verbs like *bergsteigen* ('mountaineering') and *klettern* ('climbing') are also transient collocates of *Frau* but have only been so in the last two decades.

One problem with using UFA with this query term in the data used here is certainly that the word *Frau* is overall rare and thus we face the problem of scarce data when doing statistical analysis like UFA.

Conclusion and outlook

The above statistical survey of the nomination patterns of female persons in the Alpenwort corpus revealed two things: (1) The number of references to women increase over time. Therefore, women as a group and as individuals are talked about more and this is reflected in an increase in the use of lexical genders (e.g. *Frau*) and the use of words with gender inflection ('Bergsteigerinnen'). (2) This increase occurs in connection with the increase in the number of texts specifically concerned with women in mountaineering that became apparent in the metadata (texts referring to women in the title, texts catalogued as *Frauenalpinismus*, and texts by women authors).

The most stable collocational pattern of the word *Frau*, for example, is with possessives in the phrases 'my wife' or 'his wife', which clearly indicates that they are not considered part of the in-group of mountaineers but rather in their conventional role as wives and are seen rather less as acting subjects (which is also indicated by the use of the preposition *mit* 'with' in constructions like *mit meiner Frau* 'with my wife'). This shows that the idea of the male, heterosexual, and married man writing for the Alpine Club remains stable to the present day.

Words referring to women are rarely key. If the concept of keyness is understood as referring to the 'aboutness' of texts (Gabrielatos 2018: 226), earlier texts are about

the close relationships of male authors with certain women rather than about female achievements on the mountains. If female achievements are mentioned at all, this is always in relation to men and/or highlighted as special. Only later in the corpus are women explicitly referred to as mountaineers.

This means that there is growing recognition of women in mountaineering in the texts of Alpenwort. It seems, however, to be limited to certain articles explicitly dedicated to women in mountaineering as a special topic. Therefore, writing about *Bergsteigerinnen* does not appear to be normalised in the remaining texts. This seems to be a phenomenon of second-wave feminism that also occurs in other areas: after a long time in which achievements of women were either disregarded or completely forgotten, some awareness of women's achievements was generated through the work of second-wave feminists. A reflection of this then is that women are represented as a special topic. On the other hand, it seems that even up to today women authors hardly ever write articles about other topics or more general articles that do not include self-reflexive aspects of gender, so the representation of women in the overall corpus is still rather low.

As noted above, the findings of this paper are limited to the material and specific questions (and queries) used here. As McEnery and Baker (2017: 199) argue, exploring 'marginalised identities' in corpora is a huge challenge: 'the marginal nature of the identities brings with it the strong likelihood that the frequency of the lexis referring to such behaviours and groups will be suppressed'. Thus, low overall frequencies may lead to problems with statistics and the results cannot be generalised, as was demonstrated in Rampl and Posch (2022). Therefore, it is essential to also look at the text and concordance line level. Further study of the topic should thus of course also include investigation of the broader contexts in which key items and query items are used. It would also be fruitful to make an in-depth comparison of masculine and feminine person appellations in relation to the problem of the generic masculine. One thing can be said about this in advance: If in fact, as the meta-data overview has shown, most texts are about men and also written by men, the probability that women are "mitgemeint" ('also included') and feel included by masculine generics will probably not be very high in these texts. In a controversial article in the magazine *UK Climbing*, climber and author Sarah Jane Dobner called for a discussion of the inherent sexism and colonialism in the mountaineering community (Dobner 2019). Dobner in particular states that

> contemporary anthologies really can do better, can make an active effort to recruit material from a more diverse base. If there isn't a significant proportion from non-white-male-contributors in your selected essay book – why not? Because these people have been climbing. They will have stories. (Dobner 2019)

References

Achrainer, M., F. Kaiser, F. Trojer. 2011. Berg Heil! Eine Einführung. Pp. 9–16 in *Berg Heil!: Alpenverein und Bergsteigen 1918-1945,* edited by Deutscher Alpenverein, Oesterreichischer Alpenverein, and Alpenverein Südtirol. 1st ed. Köln: Böhlau Verlag.

Aull, L. L., D. W. Brown. 2013. Fighting Words: A Corpus Analysis of Gender Representations in Sports Reportage. *Corpora* 8 (1): 27–52, https://doi.org/10.3366/cor.2013.0033.

Baker, P., C. Gabrielatos, M. KhosraviNik, M. Krzyżanowski, T. McEnery, R. Wodak. 2008. A Useful Methodological Synergy?: Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press. *Discourse & Society* 19 (3): 273–306, https://doi.org/10.1177/0957926508088962.

Baker, P. 2010. Will Ms Ever Be as Frequent as Mr? A Corpus-Based Comparison of Gendered Terms across Four Diachronic Corpora of British English. *Gender and Language* 4 (1), https://doi.org/10.1558/genl.v4i1.125.

Baker, P., T. McEnery (eds.). 2015a. *Corpora and Discourse Studies: Integrating Discourse and Corpora.* London: Palgrave Macmillan UK.

Baker, P., T. McEnery. 2015b. Introduction. Pp. 1–20 in Baker, P., T. McEnery (eds.). *Corpora and Discourse Studies: Integrating Discourse and Corpora*. London: Palgrave Macmillan UK.

Brezina, V. 2018. *Statistics in Corpus Linguistics: A Practical Guide.* Cambridge: Cambridge University Press.

Bubenhofer, N., J. Schröter. 2012. Die Alpen. Sprachgebrauchsgeschichte - Korpuslinguistik - Kulturanalyse. Pp. 263–287 in Maitz, P. (ed.). *Studia linguistica Germanica, vol. 110*, *Historische Sprachwissenschaft: Erkenntnisinteressen, Grundlagenprobleme, Desiderate; [Beiträge eines Humboldt-Kollegs ‚Wohin steuert die Historische Sportwissenschaft?' im September 2009 an der Universität Debrecen]*. Berlin: de Gruyter.

Bubenhofer, N., P. Scheurer. 2014. Warum man in die Berge geht. Das kommunikative Muster ‚Begründen' in alpinistischen Texten. Pp. 239–268 in Hauser, S., U. Kleinberger, K. S. Roth (eds.). *Musterwandel – Sortenwandel. Aktuelle Tendenzen der diachronen Text(sorten)linguistik*: Bern, Berlin: Peter Lang Ltd. International Academic Publishers.

Caldwell, D. (ed.). 2016. *The Discourse of Sport.* London: Routledge.

Dobner, S.-J. 2019. The Perfect Line: Naming and Claiming. Retrieved 5/11/2019 (https://www.ukclimbing.com/articles/features/the_perfect_line_naming_and_claiming-12053).

Gabrielatos, C., P. Baker. 2008a. Fleeing, Sneaking, Flooding. *Journal of English Linguistics* 36 (1): 5–38, https://doi.org/10.1177/0075424207311247.

Gabrielatos, C., P. Baker. 2008b. Fleeing, Sneaking, Flooding. *Journal of English Linguistics* 36 (1): 5–38, https://doi.org/10.1177/0075424207311247.

Gabrielatos, C., A. Marchi. 2012. *Keyness: Appropriate Metrics and Practical Issues.* Bologna. Retrieved 6/8/2020 (https://www.researchgate.net/publication/261708842_Keyness_Appropriate_metrics_and_practical_issues).

Gabrielatos, C. 2018. Keyness Analysis: Nature, Metrics and Techniques. Pp. 226–258 in Taylor, C., A. Marchi (eds.). *Corpus Approaches to Discourse: A Critical Review.* Oxford: Taylor and Francis.

GENDER **A** VÝZKUM

Hardie, A. 2012. CQPweb – Combining Power, Flexibility and Usability in a Corpus Analysis Tool. *International Journal of Corpus Linguistics* 17 (3): 380–409, https://doi.org/10.1075/ijcl.17.3.04har.

Hardie, A. 2014. Log Ratio - An Informal Introduction. Retrieved 27/7/2020 (http://cass.lancs.ac.uk/log-ratio-an-informal-introduction/).

Hardt-Mautner, G. 1995. Only Connect. Critical Discourse Analysis and Corpus Linguistics Retrieved 27/7/2022 (http://ucrel.lancs.ac.uk/papers/techpaper/vol6.pdf).

Hornscheidt, L. 2006. *Die sprachliche Benennung von Personen aus konstruktivistischer Sicht: Genderspezifizierung und ihre diskursive Verhandlung im heutigen Schwedisch*. Vol. 15. Berlin: de Gruyter.

Hunt, S. 2015. Representations of Gender and Agency in the Harry Potter Series. Pp. 266–284 in Baker, P., T. McEnery (eds.). *Corpora and Discourse Studies: Integrating Discourse and Corpora.* London: Palgrave Macmillan UK, https://doi.org/10.1057/9781137431738_13.

Ismail, H. 2016. Of Cover Girls and Bad Boys A Corpus Linguistic Analysis of Gendered Keywords in Malaysian Sports News Discourse. Pp. 171–189 in Caldwell, D. *The Discourse of Sport.* Oxford: Routledge.

Jaworska, S., S. Hunt. 2017. Differentiations and Intersections: A Corpus-Assisted Discourse Study of Gender Representations in the British Press Before, During and After the London Olympics 2012. *Gender and Language* 11 (3): 336–364, https://doi.org/10.1558/genl.28858.

Kilgarriff, A., D. Tugwell. 2002. Sketching Words. Pp. 125–137 in M.-H. Corréard (ed.). *Lexicography and Natural Language Processing: A Festschrift in Honour of B.S.T. Atkins*. Göteborg: EURALEX. Retrieved 3/12/2022 (https://www.kilgarriff.co.uk/Publications/2002-KilgTugwell-AtkinsFest.pdf).

Kjellmer, G. 1986. 'The Lesser Man': Observations on the Role of Women in Modern English Writings. Pp. 163–176 in Aarts, J., W. Meijs (eds.). *Corpus Linguistics II.* Leiden: Brill.

Kotthoff, H., D. Nübling. 2018. *Genderlinguistik: Eine Einführung in Sprache, Gespräch und Geschlecht*. Tübingen: Narr\Francke\Attempto.

Krendel, A., M. McGlashan, V. Koller. 2021. The Representation of Gendered Social Actors across Five Manosphere Communities on Reddit. *Corpora* 17 (2): 1–25.

Lautenschläger, S. 2016. Geschlechtsspezifische Körper- und Rollenbilder. Dissertation.

Lischinsky, A. 2018. Overlooked Text Types. From Fictional Texts to Real-World Discourses. Pp. 60–83 in Taylor, C., A. Marchi (ed.). *Corpus Approaches to Discourse: A Critical Review.* Milton: Taylor and Francis.

Lutz, R. 2002. Triumph des männlichen Willens: Helden, Kampf und Krieg im Bergsport. *Forum Wissenschaft: Sport - Politik - Bewegung* (2). Retrieved 3/12/2022 (http://www.bdwi.de/forum/archiv/archiv/441771.html).

Mautner, G. 2016. Checks and Balances: How Corpus Linguistics Can Contribute to CDA. Pp. 154–179 in Meyer, M., R. Wodak (eds.). *Methods of Critical Discourse Studies.* 3rd ed. Los Angeles, CA: Sage Publications.

McCuaig, J., H. Hartmann, M. C. Lim, C. R. Pennington, G. Reid. 2019. Positionality Map. Retrieved 3/12/2022 (https://doi.org/10.31219/osf.io/bnh7p).

McEnery, T., G. Brookes. 2022. Register, Belief and Violence: A Multidimensional Approach. Pp. 87–120 in Posch, C., G. Rampl, K. Irschara (eds.). *Wort - Satz - Korpus. Beiträge zur Korpuslinguistik*. Innsbruck: iup.

McEnery, T., A. Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

McEnery, T., H. Baker. 2017. *Corpus Linguistics and 17th-century Prostitution*. London: Bloomsbury.

McEnery, T., V. Brezina, H. Baker. 2019. Usage Fluctuation Analysis. *International Journal of Corpus Linguistics* 24 (4): 413–444, https://doi.org/10.1075/ijcl.18096.mce.

Partington, A. (ed.). 2004. *Linguistic Insights Vol. 9, Corpora and Discourse*. Bern, Berlin, Frankfurt am Main, Wien: Lang.

Partington, A. 2015. Corpus-Assisted Comparative Case Studies of Representations of the Arab World. Pp. 220–243 in Baker, P., T. McEnery (eds.). *Corpora and Discourse Studies: Integrating Discourse and Corpora*. London: Palgrave Macmillan UK.

Posch, C., G. Rampl. 2020. Lima or Cima? Structure Recognition and OCR in Building the Corpus of the Austrian Alpine Club Journal. *International Journal of Corpus Linguistics* 25 (4): 489–503.

Posch, C. 2022. Digital Linguistics: Integrating Digital Humanities, Corpus Linguistics and Critical Discourse Studies. Habilitationsschrift, Institut für Sprachwissenschaft, Universität Innsbruck. Innsbruck.

Purtschert, P. 2019. *Kolonialität und Geschlecht im 20. Jahrhundert: Eine Geschichte der weißen Schweiz*. Bielefeld, Berlin: Transcript.

Rak, J. 2007. Social Climbing on Annapurna: High-altitude Mountaineering Narratives. *ESC* 33 (1–2): 109–146.

Rampl, G., C. Posch. 2022. Methodisch geht es korpuswärts! Korpuslinguistik und Produktivität anhand des Suffixes -wärts. Pp. 119–145 in Posch, C., G. Rampl, K. Irschara. *Wort - Satz - Korpus. Multimethodische digitale Forschung in der Linguistik*. Innsbruck: iup.

Reisigl, M., R. Wodak. 2016. The Discourse-Historical Approach (DHA). Pp. 87–121 in Meyr, M., R. Wodak (eds.). *Methods of Critical Discourse Studies*. 3rd ed. Los Angeles, CA: Sage Publications.

Rheindorf, M., R. Wodak. 2019a. 'Austria First' Revisited: A Diachronic Cross-Sectional Analysis of the Gender and Body Politics of the Extreme Right. *Patterns of Prejudice* 53 (3): 302–320, https://doi.org/10.1080/0031322X.2019.1595392.

Rheindorf, M., R. Wodak. 2019b. Genre-Related Language Change: Discourse- and Corpus-Linguistic Perspectives on Austrian German 1970–2010. *Folia Linguistica* 53 (1): 125–167, https://doi.org/10.1515/flin-2019-2006.

Runggaldier, I. 2011. *Frauen im Aufstieg: Auf Spurensuche in der Alpingeschichte*. Bozen: Edition Raetia.

Secules, S., C. McCall, J. A. Mejia, C. Beebe, A. S. Masters, M. L. Sánchez-Peña, M. Svyantek. 2021. Positionality Practices and Dimensions of Impact on Equity Research: A Collaborative Inquiry and Call to the Community. *Journal of Engineering Education* 110 (1): 19–43, https://doi.org/10.1002/jee.20377.

Stefanowitsch, A. 2020. *Corpus Linguistics: A Guide to the Methodology.* Vol. 7. Berlin: Language Science Press.

Taylor, C., A. Marchi (eds.). 2018. *Corpus Approaches to Discourse: A Critical Review.* Milton: Taylor and Francis.

Tudor, A. 2011. feminismus w_orten lernen: Praktiken kritischer Ver_Ortung in feministischen Wissensproduktionen. Pp. 57–99 in *Transdisziplinäre Genderstudien, Vol.* 3, *Feminismus schreiben lernen,* edited by AK Feministische Sprachpraxis. 1st ed. Frankfurt am Main: Brandes & Apsel.

Tudor, A., L. Hornscheid. 2011. Feminismus schreiben lernen: Ein Glossar. Pp. 180–188 in AK Feministische Sprachpraxis (ed.), Feminismus schreiben lernen (Transdisziplinäre Genderstudien 3). Frankfurt am Main: Brands & Apsel.

Wodak, R. 2008. Introduction: Discourse Studies - Important Concepts and Terms. Pp. 1–29 in Wodak, R., M. Krzyżanowski (eds.). *Qualitative Discourse Analysis in the Social Sciences*. Basingstoke: Palgrave Macmillan.

Claudia Posch is Assistant Professor at the department of Linguistics at the University of Innsbruck with research interests in Feminist Linguistics as well as Corpus and Discourse Studies. She has recently completed her Habilitation project on the combination of both Corpus Linguitics and Feminist Discourse Studies. ORCID: https://orcid.org/0000-0002-3939-8306, Contact e-mail: Claudia.posch@uibk.ac.at.