





Analyzing gender clues in war-time letters

Heikki Keskustalo , Laura Korkeamäki  and Selja Vanamo
Faculty of Information Technology and Communication Sciences,
Tampere University, Finland

Kimmo Kettunen 
School of Humanities, Finnish Language and Cultural Research,
University of Eastern Finland, Finland

Sanna Kumpulainen 
Faculty of Information Technology and Communication Sciences,
Tampere University, Finland

Abstract

Many historians struggle with their information needs which cannot be directly served by the information access systems. Satisfying these needs often requires reasoning and interpretation of pieces of information in context, from user-specific viewpoints. One common need in studying historical phenomena is what indicates gender in historical text. We call such textual indicators ‘gender clues’ because they help satisfy information needs regarding the concept of gender. In this article, we analyze gender clues qualitatively and present a typology of them based on a set of private letters from the Second World War in Finland. We also discuss the general need to create metadata to support the historian’s explorations from specific viewpoints, especially in small and noisy collections that are common in the historical domain.

Correspondence:

Heikki Keskustalo, Tampere University, Tampere 33014, Finland.

E-mail:

heikki.keskustalo@tuni.fi

1 Introduction

Making sense of past information to create new knowledge is at the heart of history scholarship. Scholars often have their personal research viewpoints, which may differ from the intentions of the authors of the original texts. In the case of studying private correspondence, the historians and linguists often close-read and interpret the letters from this viewpoint. It is important to a person to find information items that contribute to the process of understanding the historical phenomena studied. Traditional information access systems typically support information retrieval at the level of the document collection. However,

supporting access into relevant pieces of information ‘within the documents’ is more challenging—especially if the goal is not topical retrieval, but to find information items that can support human reasoning (Kumpulainen *et al.*, 2020).

Historical documents are increasingly available for research purposes in digital form due to the recent advances in, e.g. optical text recognition (see Muehlberger *et al.*, 2019). However, despite the efforts, many scholars still struggle with digital practices, e.g. in expressing their versatile information needs to the information systems (Korkeamäki and Kumpulainen, 2019) or due to difficulties derived from multiple contextual reasons such as socio-

organizational, collection-related, and research task-related problems (Kumpulainen and Late, 2021). Furthermore, while systems are efficient in providing access to documents, access into relevant information items ‘within’ the documents is still challenging. Developing this access requires that both the user needs are understood—and the reasoning related to relevant information items. We shed light on this complex issue by focusing on one specific case. We will inspect ‘textual gender clues’ in private Second World War letters to explain how they allow making claims about the gender of the recipients and senders of the letter.

Private letters constitute an important source type in historical research. They offer valuable complementary data for understanding historical events. For example, during the war time accidental shootings are not necessarily recorded in the official war diaries of the troops—but private correspondence may bring light into such events (Nurminen, 2015). Due to their personal nature, private letters empower researchers to ask specific type of questions, such as how human emotions are expressed during the times of crisis, and how the relationship between the persons in dialogue affects the discourse (Hagelstam, 2011; Taskinen, 2015; Mustanoja, 2017).

‘Gender issues’ gain attention in many research areas nowadays (see, e.g. Nielsen *et al.*, 2017). They are also important when the historians and linguists study historical correspondence. Examples of such studies include asking how the personal relationship affects the subjects kept silent about between a soldier and his wife, or between brothers-in-arms (see, e.g. Hagelstam, 2011; Hagelstam and Taskinen, 2015); how the ethos of ‘home, religion, and fatherland’ was expressed in letters sent between the war front and the home front (Mikkola, 2017); and how men expressed emotions and maintained their friendship (Riihonen, 2017). Other examples include the topics of argument between a married couple (Pajunen, 2017); discourses of couples belonging to different age groups (Silén, 2017); and discourses between a mother and young daughter (Ponkala, 2017). The prevalence of information needs related to gender—and the lack of automated support for them—inspired us to focus on gender expressions in the context of classifying/retrieving them.¹

During historical research, tens of letters—sometimes even hundreds—are typically close-read and interpreted by the historian. From this viewpoint, we define a ‘clue’ as a string which can be interpreted as evidence for a task-specific need. Words, phrases, and fractions of words can act as clues. In case of gender clues, the ‘clue status’ of a string is volatile and depends on the reader’s intentions and the context (cf. Pirolli and Russell, 2011). Clues resemble ‘concept symbols’ which constitute associative links from citations to concepts, as discussed by Small (1978). In our case, a gender clue proposes that a gender concept is possibly present. In letters, both the position and context affect the meaning and interpretation of the string. For example, due to the letter writing conventions, the beginning of a letter often informs about the recipient (e.g. opening greeting ‘Dear Mary’). In this study, we focus on the greetings and signatures of letters (i.e. opening and closing salutations). While we focus on one language and data type—Finnish war-time letters—our observations are largely language-agnostic.

We have three research questions: (RQ1) What type of gender clues exist in the greetings and signatures of letters? (RQ2) What kind of reasoning allows observing them? (RQ3) Could we facilitate automatic observation of these clues? Regarding the first and second questions, we focus on both the first names and ‘other kinds of gender clues’ we observed. Regarding the third question, we explain the types of clues observed and whether they could be spotted automatically. Our emphasis is on the qualitative analysis of the types of clues observed.

As recently pointed out by Oberbichler *et al.* (2021), interdisciplinary research in the historical domain may involve various steps, such as engineering and manual annotation, which are necessary but may be scientifically uninteresting. We also observed the necessity to perform various complementary steps to reach our research objectives, while simultaneously negotiating how much information—and how deeply—we could process at each step. At the first step, we manually added field tags into the originally unstructured digitized letters to annotate the greetings and signatures in them. Secondly, we realized a simple computer program to allow the automatic retrieval of letters based on selected text fields and word-level gender metadata (male or female) which was

automatically available. The program allowed retrieving letters based on the gender of the recipient and sender as the search criterion (e.g. retrieve a letter written to a female by a male). We utilized the open-source morphological analyzer ‘Omorfi’² (see Pirinen, 2015) to associate individual words with gender metadata. As the third step, we analyzed all gender clues in the greetings and signatures intellectually to understand the reasons for successful and failed retrieval. The gold standard information regarding the actual gender of the author and recipient was available produced by an academic historian for his own research purposes. Our motivation for performing the steps was to increase understanding regarding issue of how conceptual access into gender could be supported automatically in the future. The steps described above, comprising mixed methods, were required to approach this question.

In the next section, we start by discussing the concept of gender from the point of view of the language. Section 3 describes the data and the analysis method. In Section 4, we present the findings, and Sections 5 and 6 present the discussion and conclusions regarding the prospects of supporting conceptual access to gender clues automatically.

2 Gender in Language

While languages can express gender in many ways, usually four conceptualizations are distinguished, namely ‘grammatical’, ‘lexical’, ‘referential’, and ‘social’ gender (Hellinger and Bußmann, 2001). Grammatical gender is an inherent property of the noun, and it determines the agreement between the noun and its satellite element (e.g. a pronoun or an adjective). Typically, gender languages have two or three gender classes—often including ‘feminine’ and ‘masculine’. A language may also reduce the number of grammatical gender classes or even lose the original gender system completely—as in the case of e.g. English (Hellinger and Bußmann, 2001). Lexical gender relates to the property of extralinguistic femaleness or maleness such as personal nouns like ‘sister’ and ‘brother’. The third conceptualization, referential gender, relates linguistic expressions to the non-linguistic reality. It identifies the referent as male, female, or gender-indefinite. German expression *das Mädchen*

(‘a girl’), for example has lexical-semantic specification as ‘female’ and is generally used to refer to females (Hellinger and Bußmann, 2001). Last, social gender is related to conventionalized ways of thinking. Speakers of language may tend to think referents of certain words as female or male, although both interpretations are possible and the word itself is gender neutral. Examples of this include e.g. certain terms of profession like ‘lawyer’, ‘surgeon’, which are easily thought of as male (Hellinger and Bußmann, 2001; Laakso, 2005).

Languages express gender in different degree: some code gender in their linguistic structure and form more, and some less. Finnish—the language of our case study data—lacks grammatical gender, as do all Finno-Ugric languages (Laakso, 2005). There are two feminine suffixes in Finnish derivational morphology (-tAR and -kkO), which show the femaleness of the word.³ The third person pronoun ‘hän’ (he/she) is the same for everyone, thus its usage does not propose about the gender. The absence of grammatical gender has contributed to the perception of Finnish as a comparatively gender-neutral language (see Laakso, 2005, pp. 101–109). While this is true to some extent, the so-called male bias (male gender is considered as a neutral and prototype gender) can be found in Finnish. For example, occupational terms have used so far mostly masculine terms.⁴ Regarding the use of such terms, the situation is changing, but in the era of our data (1939–45) masculine terms were accepted for use.

‘First names’ in Finnish can be divided to large extent to male and female names.⁵ Gender distinction is part of the inherent lexical knowledge of these names. In the era of our historical letters some first names do not make gender distinction while still showing a strong bias for either sex. This kind of ambiguity is rare in Finnish (Laakso, 2005).⁶

‘Personal nouns’ showing close personal or familial relationships may signal about gender, e.g. brother, sister, ‘aunt’, ‘uncle’, and ‘wife’. In Finnish such words (*veli, sisko, täti, setä, vaimo*, etc.) are lexically gender marked, while in many other languages the gender could be also marked grammatically (as *der Bruder* in German).

‘Titles’ are common in the openings and closings of letters, e.g. ‘Mr’, ‘Ms’, ‘private’, ‘sergeant’, ‘teacher’, but not necessarily used by the authors of letters. Some titles

carry lexical gender e.g. soldiers of the 1940s were male. In Finnish, military rank *sotamies* ('private', 'infantryman') explicitly states this via ending part *mies* ('man'). Considering the Finnish society of the 1940s, lexically gender-neutral titles such as teacher may show gender bias (in this case toward male). In our analyses, we will focus on the various types of expressions—beyond the categories described above—which might help propose the presence of female or male gender.

3 Data and Experimental Setup

3.1 War-time letters

We focus on the private correspondence between persons in the war front (soldiers) and home front (civilians) in Finland during the World War II (1939–45). Sending letters was the only way to keep regular contact between the war and home fronts. Approximately 1.14 billion letters, postcards, and parcels were delivered via Finnish Army Field Post during World War II in Finland (Taskinen, 2021). Our data consist of a sample of 3,094 digitized letters, originally created via optical character recognition (OCR) of typeset transcriptions of initially handwritten letters. It is a subset of larger Wartime Letter Collection available at the Tampere University Folklife Archives, Finland (Kper) (see Taskinen, 2021, pp. 375–384).

Finnish became established as a written language during the 19th century. The ongoing effort to standardize the written expressions can be observed, for example in the newspapers of the era containing alternative spellings and coexisting inflectional variants (Järvelin *et al.*, 2015). The private letters of the war time (1940s), on the other hand, often portray casual face-to-face communications. As the authors of the letters may have been inexperienced writers, the letters often contain grammatical errors, lack capital letters or punctuation marks, and may contain oral phrases and dialect in written form (see Taskinen, 2021, pp. 176–185).

War-time letters are challenging from the IR point of view for many reasons. The letters do not have explicit structure, differently from the modern emails having an automatically enforced structure including, e.g. fields for the receiver and sender. Additionally, from the natural language processing point of view, our target letters have many challenges. First, Finnish is highly inflectional, in which words are written

together to form compound words, and suffixes are added to word stems to express inflection and derivation (Alkula, 2000). This leads to a potentially huge number of word forms that could exist. Secondly, due to the errors caused by the OCR process, noisy words occur, containing misinterpreted characters. Third, the dialectical and historical expressions are prevalent in personal letters written often in informal style. Fourth, the authors of the letters did not necessarily follow the letter writing conventions, such as even using always greetings and salutations, and the personal writing styles and skills of authors vary (cf. Taskinen, 2021, pp. 178–181). These aspects occur simultaneously, leading to, e.g. occurrences of words which are noisy inflectional variants of historical dialect. To connect such words with non-topical information needs is non-trivial.

In this study, we limit our attention to the 'greetings' (to initiate the letter) and the 'signatures' (to sign off). The other parts of the letter obviously may allow reasoning about various gender-related issues, e.g. about the gender of the sender, recipient, and the target persons discussed. See Taskinen (2021, pp. 46–85) for examples of the kinds of research questions that can be addressed via close reading historical letters. Waldvogel's (2007) observations regarding the greetings of the modern workplace emails are relevant also in the historical context of the present study. Greetings and signatures allow personalization of the messages, expressing affection, and setting the general tone for conversation. They contribute to the construction and maintenance of relationships between people. As a personal linguistic preference, they also encode and construct social information—status, distance, and gender identity. The overall context, e.g. general culture and organization, also affects their use (Waldvogel, 2007). For example, the use of endearments in correspondence between a close couple, or the friendly tone of salutations between men serving together, may help the human observer indicate the type of relationship they had.

3.2 Manual annotation of salutations

Kumpulainen *et al.* (2020) reported three important person roles in letters based on historians interviewed: the recipient, the sender, and the target person(s) discussed. We hypothesized that the greeting and signature offer a natural starting point for studying the

possibilities of recognizing the gender of the recipient and sender—and finding the gender–gender ‘direction’ (e.g. ‘letter to a female from a male’). The remainder of the letter, besides greeting and signature, obviously may inform many things regarding the gender issues for the human interpreter—including the gender of the sender, recipient, and target person(s) discussed. However, in this study we limit our attention to the greetings and signatures.

We first manually marked up the greetings and signatures in the (originally unstructured and unmarked) letters to prepare the test data. We started by familiarizing ourselves with a small subset of letters to develop rules for detecting greetings and signatures (Tables 1 and 2, respectively).

One of the authors annotated the letters, while the researchers discussed the proper practices of annotation during the process. After having gained the first-hand experience regarding the problems encountered and effort required, we continued by annotating a larger set. At this point, we decided to mark up the boundaries between the letters; the beginning and ending points of greetings and signatures; and the date (if such information was available). As the last step, a larger subset of letters was annotated, during which we were able to utilize existing file name coding automatically, which expressed the date and sub-collection numbering information. Table 3 shows an original Finnish sample letter (on the left) with the XML mark-up produced, with English translation (on the right). Importantly, greetings (openings) were manually marked up as *bsal* field (‘beginning salutation’) and signatures as *esal* field (‘ending salutation’), whenever they were observed. Both types of salutations occur within the ‘text’ field. The contents of salutation fields were analyzed at a later, separate phase, by a different researcher. In that phase the point of view was changed from recognizing salutations to observing and reasoning about possible gender clues present in the salutations.

3.3 Automatic retrieval

Our automatic gender-based retrieval was based on the idea that the recipient is typically addressed in the beginning of the letter, and the author/sender at the end. We designed a simple retrieval program in Python utilizing gender-specific words observed in

Table 1. Manual annotation rules for greetings (opening salutations)

Description of the rule
Opening statement, and name or name-like expression.
Opening statement without name or name-like expression.
Greeting is missing.

Table 2. Manual annotation rules for signatures (closing salutations)

Rule description
Ending statement and name/name-like expression with one to two other words in the same sentence.
Ending statement and name/name-like expression with more than two other words in the same sentence.
Ending statement and name or name-like expression with more than one sentence.
Only name/name-like expression.
Greeting without name/name-like expression.
Signature is missing.

greeting and signature fields (Fig. 1). Gender was observed based on metadata (female or male) of morphological analyzer Omorfi which allowed recognizing it based on the first names of persons.¹

We retrieved the letters using four gender–gender combinations (female, female), (female, male), (male, female), and (male, male) as the search criteria. The first and second items of the pair denote, in order, the genders of the ‘recipient’ and the ‘sender’. For example, (female, male) denotes letters to female recipient by a male sender. Our retrieval method successfully retrieved such a letter if it contained, e.g. the opening greeting ‘Dear Mary’ and signature ‘Best regards, Mark’. The word ‘Mary’ is automatically associated with ‘female’ metadata, and ‘Mark’ with ‘male’. The baseline truth regarding the actual gender–gender combinations ($N = 3,094$) was available separately, created by a professional historian. This metadata was sometimes incomplete ($N = 42$), but for most of the letters ($N = 3,052$) the precise gender–gender combination was known. The baseline truth was used as the gold standard when we analyzed the gender clues. The gender metadata of the morphological analyzer was based on the first names only. Therefore, we paid attention in our intellectual

Table 3. Letter sample (pseudonymized) showing the tag structure added to the unstructured historical letters

Data sample	Approximate translation
<letter>	<letter>
<globalid > 650</globalid>	<globalid > 650</globalid>
<collection > 108</collection>	<collection > 108</collection>
<localid > 10</localid>	<localid > 10</localid>
<date > 19430909</date>	<date > 19430909</date>
<comment>NULL</comment>	<comment>NULL</comment>
<text>	<text>
... kirje: SAK/105/11	... letter: SAK/105/11
Täällä Samassa paikassa 9.9.43.	Here in the same place 9.9.43.
<bsal>Rakas Vaimoni</bsal>	<bsal>My Dear Wife</bsal>
Monet lämpimät terveiset sinulle ja kiitos kirjeistä ...	Many warm greetings to you and thank-you for the letters ...
<esal>jos taas lopetan tämänkin juuri ja	<esal>if I just finish this again and
Rakas Liisa voi taas muuta ole	dear Liisa can not otherwise be
tervehditty ja hyvästi juhosi</esal>	greeted and good-bye your juho</esal>
Toisille myös terveiset. Reunamerkintä: ...	Regards to the others also. Marginal note: ...
</text>	</text>
</letter>	</letter>
<globalid>	<globalid>

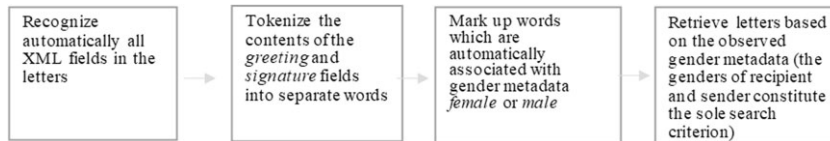


Fig. 1. Retrieval of historical letters based on the gender of the recipient and sender

analyses on all ‘other’ types of gender clues to learn about the conceptual access points (via gender clues) possibly present in the texts.

3.4 Intellectual analysis of clues

We analyzed gender clues in three situations: (1) when relevant letters were missed in retrieval, (2) when relevant documents were successfully retrieved, and (3) when non-relevant documents were retrieved. The gender clues were coded in three rounds by one researcher. The coding process started by getting familiar with the research data (cf. [Cohen et al., 2018](#), p. 669). The main goals for the first round were understanding what kind of search results were found or missed, and what kind of ‘gender clues’ appeared in the salutations, as well as developing a list of initial codes for gender clue types. The researcher made notes of possible codes and provided examples of the salutations. Some codes for gender clues, such as ‘kinship’ and ‘military rank’, were derived from [Kumpulainen et al. \(2020\)](#). However, most of the

codes were derived intuitively from the data. Codes were also created for the causes of errors in retrieval of the first names. One obvious cause was the presence of OCR errors. For other causes of errors, initially longer phrases were used to describe the problems in the search process. In this round the codes and examples were written to a Word document, allowing, e.g. describing freely the errors in the search process using one’s own words, longer phrases, and more descriptive language. The focus of the first round was mainly in describing the phenomenon and collecting a list of interesting cases that could be later used as codes. At the end of the first round (and preparing for the second round), a rehearsal coding was done to a small part of the research data, using an Excel spreadsheet. Salutations were coded to test the usability of the initial codes. At this point it was clear that the longer phrases needed to be shortened, to make the codes usable in Excel. The list of initial codes and examples of salutations were discussed in a group of three researchers, and the code ‘endearment’ was added to

the coding scheme. The importance of marking all difficult or unclear cases in salutations was also discussed. Moreover, we decided to add a ‘notes’ column to the spreadsheet to allow free comments about the salutations.

The second round of coding aimed at systematic coding of the beginning and ending salutations of the four gender–gender combinations, by using the codebook from the first round. The names and definitions of the codes, as well as the groupings of the codes to different columns, evolved during the second round. Mainly, a single Excel cell was used per code. Creating definitions for the codes was a challenge, partly because of the four distinct gender–gender combinations in the research data, and this change of perspective needed to be kept in mind during the coding. The three types of retrieval situations were studied further in the following way.

Relevant letters were ‘missed’ during retrieval, because either there were no first names in the salutations, or the first name was not automatically recognized by the search engine. These cases were labelled with codes such as ‘first name is missing’, ‘first name not found’, or ‘salutation is missing’. It was most interesting to know what kind of ‘other gender clues’ (except first names) could be found in the salutations. For this purpose, codes for person-associated gender attributes were created (‘endearment’, ‘kinship’, ‘marital relationship’, ‘military rank’, ‘nickname’, ‘title expression’, and ‘other’ gender-related term). Last, a code for such salutations in which the gender-related clue was completely missing was created (‘gender-related clue is missing’).

Both in case of ‘relevant’ and ‘non-relevant’ letters ‘retrieved’, it was important to understand why such search results were found. For example, relevant letters were sometimes retrieved but on the wrong basis. Therefore, a specific code ‘correct hit on the wrong basis’ was developed for this type of search results. The code was used together with other codes to elaborate the reasons for retrieving relevant result but on the wrong basis. Interestingly, both relevant and non-relevant letters were sometimes found, based on first name, but going against our main assumption. Surprisingly, sometimes the receiver’s name occurred in the ending salutation, or the sender’s name in the beginning salutation or some other first names. The corresponding error codes were used for these cases.

Last, in some cases the search engine found a gender clue based on a word which was actually not a first name; in these cases the code ‘word incorrectly retrieved as first name’ was used.

The third round was needed to secure the consistent use of the codes. For example, words ‘sisko’ (sister) and ‘veli’ (brother) were coded as ‘kinship’, although in capitalized form they are also first names in Finnish. The researcher checked the codes used in previous round, and made corrections if necessary.

4 Results

4.1 Findings

We next describe the gender clues observed in three cases: when the retrieval failed to retrieve relevant documents; when retrieval was successful; and when non-relevant documents were retrieved. We consider the fourth case to be uninteresting—when non-relevant documents were not retrieved—and did not analyze this case. Specifically, in this type of analysis, the gender clues in *all* relevant letters were inspected regarding all four gender–gender combinations. Our purpose is to explain the grounds for developing gender-based retrieval in the future. One might expect that our proper name-based retrieval leads to high precision and low recall, because gender clues are detected using a very narrow but often precise criterion, while the proper names may be missing from individual letters. Our results demonstrate the insufficiency of proper name-based gender metadata as a sufficient source of gender information (Table 4).

Table 4 indicates that a high precision was observed in gender directions to male from female (MF) and to female from male (FM) and low in directions to female from female (FF) and to male from male (MM), where a larger share of non-relevant documents was retrieved. The recall was moderately low for all gender directions. These mixed results were due to a great variability of actual gender expressions in letters—which was only partially captured via our proper name-based searching. Next, we will dig deeper into this problem and describe the actual gender clues observed in the letters.

Table 4. Effectiveness of gender-based retrieval of historical letters

Effectiveness	Gender direction (gender of recipient followed by gender of sender)			
	to FF	to FM	to MF	to MM
Precision	42.1%	96.1%	87.0%	38.7%
Recall	25.0%	18.5%	30.3%	22.4%
Relevant retrieved	85	366	174	36
Relevant missed	255	1,610	401	125
All relevant	340	1,976	575	161
Non-relevant retrieved	117	15	26	57

4.2 Relevant documents missed

We start by describing the gender clues present in relevant documents, which our retrieval method failed to retrieve ($N=2,391$). These ‘potential’ clues are especially important to understand. Table 5 presents a classification of different types of gender clues observed in the greetings and signatures. Colour coding is used to show how common each clue type was among the combination of the salutation type (greeting or signature) and direction (recipient and sender). The darker shades of grey in each vertical column indicate that the clue type was more common relative to the lighter shades. For example, the seventh column in the heat map (‘Signatures/MF’) indicates that in the signatures of letters to male recipients written by female authors, military titles were missing (lightest shade of gray in cell). The female authors did not possess military titles. Moreover, kinship and marital expressions were the most common gender clues (the darkest shade).

The proper names were often missing from the salutations or could not be detected automatically, thus relevant documents were missed. The salutation itself could be missing, in some cases possibly due to the writer’s unfamiliarity with the letter writing conventions (cf. Taskinen, 2015, pp. 15–16). Moreover, OCR errors caused the automatic gender recognition to fail, even though the gender was apparent to a human assessor.⁷ Other missed cases included gender clues via homographic expressions.⁸ Moreover, proper names in the signatures did not always refer to the sender.⁹ Last, showing the challenging nature of the retrieval task, one signature ended with expression *Good night Eino*—thus referring to the (male) recipient—not to the sender as expected.

Regarding the other gender clues than proper names, first, ‘kinship terms’ were common in the

openings and closings of letters (see Table 5). Female clues included expressions referring to ‘mother’, ‘sister’, ‘little sister’, and male clues to ‘father’, ‘uncle’, and ‘brother’, implying a male recipient.¹⁰ Correspondingly, kinship expressions in signatures, e.g. ‘little mother’, ‘Grandmother’, ‘father’, and ‘brother’ implied the gender of the sender.¹¹ The expressions entailed string-level variation due to the inflection, dialect, capitalized initial letter, and noise due to OCR errors. Sometimes multiple genders were implied, e.g. as in signature by ‘mother and father’.

Secondly, ‘nicknames’ were common in greetings and signatures. They often consist of informal name-like expressions, which are made-up or may concur with actual proper names. They may be gender specific, but without knowing their origin it may be impossible to deduce the gender based on the name alone.¹² Nicknames included inflectional and noisy variants, making the automatic recognition challenging. Suffixes indicating possession in nicknames can be interpreted to express closeness of relationship.¹³

Third, ‘titles’ and ‘military ranks’ were common in greetings and signatures. Some female titles are common in the general historical context (early 1900s) but are rarely used in modern language.¹⁴ Other titles were common in the specific historical context.¹⁵ Last, some titles are generally gender specific, such as ‘Rouva’ (Madam) and ‘neiti’ (miss). Regarding the male clues, ‘military ranks’ express gender, because Finnish soldiers were male. General gender-specific expressions include ‘Herra’ (Mister) often used in letters. The expressions of title and military ranks also included noisy expressions and abbreviations.¹⁶

Fourth, ‘marital expressions’ were common in salutations. This clue type occurred only between the gender directions FM and MF (see Table 5).

context). Unsurprisingly, marital expressions also included inflectional, noisy, and dialectical variants.

Fifth, gender-specific ‘endearments’ were found in all gender directions except MM. Affectionate variants of ‘proper names’ and ‘kinship terms’ constituted both female¹⁹ and male clues.²⁰ The ‘possession’ (marked by suffix *-si*) helped signal about the closeness of the relationship.²¹

Other types of clues were also observed. Gender may be apparent for the human interpreter in compound words such as ‘Sotasisareni’ (my war sister), and ‘aseve 1 i’ (noisy form of ‘brother in arms’) based on the component words ‘sisar’ (sister) and ‘veli’ (brother). This kind of reasoning could be mimicked automatically by observing gender-specific components as sub-strings of compound words written.

Last, gender clues were often missing. Greetings and/or signatures sometimes indicated a single person or a group but without gender information.²² Sometimes initial letter was used to refer to a person, like in signature *Goodbye. K.* Additional information is required to disambiguate gender in this case.

4.3 Relevant documents retrieved

When relevant documents were successfully retrieved, the gender of the recipient and sender was typically correctly indicated based on the first names. However, relevant documents were sometimes retrieved for the wrong reasons. The most common reason across all gender combinations was that the greeting or signature contained multiple names. When this occurs in the greeting (referring to recipients) or signature (referring to senders), it is non-trivial to select the correct gender direction as indicated by the gold standard information.

Homographic expressions were the second most common reason for accidentally retrieving relevant documents. For example, the frequently used idiomatic expression ‘With many [regards] ...’ was automatically connected to a female name.²³ The interpretation is theoretically correct, but unreasonable in practice. Therefore, in gender directions FF and MF (requiring a female sender), this expression facilitated retrieving relevant documents although for the wrong reason, but not in directions FM and MM where the male sender was required. A person name appearing in the signature—not referring to the

sender—facilitated retrieving relevant documents for the wrong reason in some cases.²⁴

Last, gender recognition was sensitive to OCR noise obfuscating the correct form of the proper names. Still, relevant documents could be found in some cases, although for the wrong reason.²⁵

4.4 Non-relevant documents retrieved

The most common reasons for retrieving false drops, observed in all gender directions, were that the salutations contained ‘multiple names’ belonging to different genders, and that the ‘proper name’ in the signature did not refer to the sender.²⁶ In one case, a letter started with the Biblical metaphor ‘Eevalle Aatamilta’ (to Eve from Adam). As also the signature contained a male name, the letter was incorrectly retrieved based on the search criterion ‘to a male from a male’. For the human reader it seems obvious that the letter is written to a woman by a man (the case of FM).

Homographic expressions were also an important reason for false drops. Idiomatic salutations contain homographic words that theoretically could be interpreted as female or male proper names.²⁷ Last, also in case of false drops the gender recognition was sensitive to OCR noise.²⁸

5 Discussion

Both the ‘type’ of intellectual reasoning and its experienced ‘certainty’ varied when greetings and signatures were inspected. In many situations, the gender of the recipient or sender could be inferred with great certainty, while in other cases it was obvious that the clue is missing—or that the clue was uncertain or ambiguous. Regarding RQ1, the main outcome of our analyses was a classification of gender clues into categories—‘kinship expressions’, ‘nicknames’, ‘titles’, ‘military ranks’, and ‘marital expressions’—in addition to proper names of persons. Regarding RQ2, human reasoning may utilize these categories to deduce the gender of the recipient (greetings) or the sender (signatures)—or both. Mapping of textual clues observed in the letters into these clue categories was often possible for the human regardless of noisy and inflected words, and improper tokenization. Importantly, we discovered that the word inflection may be essential in reasoning. As an example,

Table 6. Problems and solutions of automatic recognition of gender clues

Problems	Solutions
Proper name was missed because it was <i>missing</i> from the lexicon	Augment lexicons with historical proper names combined with gender-specific metadata
Proper name was missed due to OCR <i>noise</i>	Apply fuzzy matching to suggest clue candidates from lexicon containing historical proper names
Other types of gender clues systematically missed (types: <i>kinship</i> , <i>nickname</i> , <i>title</i> , <i>military rank</i> , and <i>marital relationship</i>)	Augment lexicons with new <i>clue types</i> and metadata including words pertaining to the historical context
Gender clue of type <i>endearment</i> was missed due to the missing endearments with metadata; the role of <i>possessive expressions</i> was not utilized	Construct historical endearment lexicons with gender metadata; propose close relationship based on detecting possessive suffixes
Gender clue missed, because <i>component word</i> was not analysed	Apply fuzzy matching to suggest gender-specific sub-word expressions
Gender clue observed in homographic expression in idiomatic salutation caused a false drop	Apply stop word lists to avoid the most common misinterpretations

expression ‘Aatamilta’ (from Adam) in the opening greeting is sufficient to deduce a male sender via suffix *-lta* expressing ‘ablative’ even though opening greeting usually refers to the recipient.

Our automatic recognition of gender was based on the metadata of proper names (female/male). Different problems of this approach are explicated in Table 6 (left column), together with proposed solutions to facilitate better recognition (right column).

Regarding RQ3, the system design should take advantage of our lessons. Qualitative approaches and automatic detection should be combined. Even a limited set of clue words allows conceptualizing (generalizing) them for a human reader as representatives of an abstract *clue type* (e.g. ‘military rank’), which can be observed during close reading. To support automated discovery of these clues, the corresponding expressions (e.g. ‘sergeant’, etc.) could be collected systematically with relevant metadata to form historical clue lexicons. Clue discovery could be automated by matching the document expressions with the entries of clue lexicons via fuzzy string matching and suggesting the matching expressions as clues for the end user.

Järvelin *et al.* (2015) illustrates the complexity of historical expressions. At the level of words, variant expressions include historical spelling variants, abbreviations, typesetting errors, inflectional variants, noisy variants due to OCR errors, and their combinations. Moreover, historical text may contain semantically related expressions such as synonyms, broader and narrower terms, and parts of compound words (Järvelin *et al.*, 2015). In small and noisy collections

this considerable string-level variance challenges the automated learning of different types of clues, while a human interpreter may recognize previously unseen types of expressions as obvious clues, e.g. a rare, noisy expression of a military title serving as a (male) gender clue. The challenges of connecting lexical entries with inflected and noisy target words need to be addressed during matching process (for solutions, see, e.g. Järvelin *et al.*, 2015). Also the word endings need to be notified as potentially valuable clues, because suffixes in Finnish may signal about the ‘closeness of relationship’ (via expressing possession or diminutive form) and ‘direction’ between the addressee and the addressor (via forms of ‘ablative’ and ‘allative’). ‘Stop word lists’ should be developed to allow questioning the nonsense interpretations of homographic expressions appearing as part of idiomatic salutations, as the clues proposed should reflect the user’s intentions, not rigid linguistic completeness for its own sake. Last, clue typologies are probably viable also in situations where the ‘machine-learning’ approaches are not practical due to small target text collections and unavailability of annotated training data, which is common in the historical domain.

6 Conclusion

Due to many large-scale historical digitization projects, vast amounts of historical documents are now available for researchers. Historians still often struggle with information needs which cannot be served

directly by the present information access systems. The point of departure of this article was the observation that *gender-specific access points* are often important for historians studying private correspondence. It is challenging to support this information need at the system level in digital letter collections. To understand the phenomenon, we studied the manifestations of gender clues in the greetings and signatures of historical letters. Our qualitative analysis revealed that:

- (1) ‘The whole words’ may inform about the presence of gender in different ways—especially based on words belonging to specific gender clue types (see Table 6).
- (2) ‘The word endings’ (suffixes) in salutations in a morphologically complex language (Finnish) may inform about the closeness of relationship via possession (‘-ni’, ‘-si’) (my, yours) or direction regarding the sender and the recipient (‘-lle’, ‘-lta’) (to, from). In the morphologically simpler languages, these aspects can be expressed via multi-word phrases.
- (3) ‘The component words’ of compounds written together (e.g. ‘voimamies’—strong man) may inform about the gender of the person addressed.

We observed that the metadata typically available via off-the-self linguistic tools—valuable in many respects—may not support the recognition of specific cognitive access points. Still, the human interpretation (via close reading and reasoning) may allow recognizing them. Our analysis revealed ‘clue categories’, which could be utilized in the future by annotating related textual expressions with selected metadata. The categories (e.g. kinship) and relevant expressions (e.g. brother) need to be gathered first to form *clue lexicons* sensitive to the historical context.

In a morphologically complex language, such as Finnish, word inflection may carry important information revealed for the human interpreter during close reading. Yet, word-level lemmatization is often the norm in building search systems. If the goal is to support task-specific access via clues, the norm of word-level lemmatization should be questioned, as it may eliminate valuable clue information. In English (a phrase-oriented language), similar clues may manifest as ‘multi-word phrases’ instead ‘**to my** wife’ (vaimol^{eni}). Regarding the compound words (written together), decomposition is needed to reveal the

embedded clues. In English, compounds are written as phrases, thus decomposition can be avoided. While our observations point a way toward the study of gender clues in other languages, we leave this for the future studies.

Engerer (2021) has pointed out that people not only retrieve documents for known needs, but they also interact with systems in an exploratory way. Historians can find textual clues to relevant concepts based on close reading the documents combined with human reasoning. Our observations indicate that automated support for finding gender clues would benefit from building specific-purpose linguistic tools and questioning the norm of word-level lemmatization. In the future, we intend to investigate the feasibility of offering support for historians’ cognitive access points in a naturalistic setting, facing the challenges of noisy data and complex information needs of the users.

Acknowledgements

We thank Dr Ilari Taskinen for useful comments and providing us the letters and the relevance data, and FIRE research group for useful comments, especially Dr Elina Late. We also thank Dr Boyang Zhang for working with the data in the early phases of this research.

Funding

This work was supported by the Academy of Finland grant number 326616.

References

- Alkula, R. (2000). *From Plain Character Strings to Meaningful Words*. Doctoral Thesis, University of Tampere, Tampere. Acta Electronica Universitatis Tamperensis, 51 [in Finnish].
- Cohen, L., Manion, L., and Morrison, K. (2018). *Research Methods in Education*, 8th edn. London: Routledge, Taylor and Francis Group.
- Engerer, V. P. (2021). ‘Attention, attention, exploring minds acknowledge digital structure!’ The shift to digital humanities has happened, so what should information

- scientists do in response? *Digital Scholarship in the Humanities*, 36(1): 43–63.
- Hagelstam, S.** (2011). Rakas sotilaani! Kirjeiden kirjoittaminen jatkosodan poikkeusoloissa. In Leskelä-Kärki, M., Lahtinen, A. and Vainio-Korhonen, K. (eds), *Kirjeet ja Historiantutkimus*, Historiallinen Arkisto 134, SKS, Helsinki, pp. 299–326.
- Hellinger, M. and Bußmann, H.** (2001). Gender across languages. In Hellinger, M. and Bußmann, H. (eds), *Gender across Languages: The Linguistic Representation of Women and Men*, Vol. 1. Amsterdam: Benjamins, pp. 1–27.
- Hagelstam, S. and Taskinen, I.** (2015). Hiljaisuuden myytit. Postisensuuri, vaikeneminen ja etulinjan kokemukset sota-ajan kirjeissä. In Tikka, M., Taskinen, I. and Nevala-Nurmi, S.-L. (eds), *Kirjeitä Sodasta: Kirjoittamisen Tavot ja Merkitykset Kriisiaikoina*. Postimuseo ja Tampereen Historiallinen Seura, Tampere, pp. 44–59.
- Järvelin, A., Keskustalo, H., Sormunen, E., Saastamoinen, M., and Kettunen, K.** (2015). Information retrieval from historical newspaper collections in highly inflectional languages: a query expansion approach. *Journal of the Association for Information Science and Technology (JASIST)*, 67(12): 1–19.
- Korkeamäki, L. and Kumpulainen, S.** (2019). Interacting with digital documents: a real life study of historians' task processes, actions and goals. In Azzopardi, L., Halvey, M., and Ruthven, I. (eds), *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. New York: ACM, pp. 35–43.
- Kper.** Wartime Letter Collection at Tampere University Folklife Archives, Tampere, Finland.
- Kumpulainen, S., Keskustalo, H., Zhang, B., and Stefanidis, K.** (2020). Historical reasoning in authentic research tasks: mapping cognitive and document spaces. *Journal of the Association for Information Science and Technology*, 71: 230–41.
- Kumpulainen, S. and Late, E.** (2021). Struggling with digitized historical newspapers: contextual barriers to information interaction in history research activities. *Journal of the Association for Information Science and Technology*, 1–13. Advanced Access published in November 2021. <https://doi.org/10.1002/asi.24608>.
- Laakso, J.** (2005). *Our Otherness: Finno-Ugrian Approaches to Women's Studies, or Vice Versa*. Wien: LIT.
- Mikkola, M.-R.** (2017). Koti, uskonto ja isänmaa; sodanjohdon eetos sotilaiden ja kotirintaman kirjeissä. In Mustanoja, L. (ed.), *Arjen Sirpaleita ja Suuria Tunteita: Kirjeet Sodan Sanoittajina ja Ihmissuhteiden Ylläpitäjänä 1939–1944*. Tampere Studies in Language, Translation and Literature: B5. Tampere, Tampereen Yliopisto, pp. 21–39.
- Muehlberger, G., Seaward, L., Terras, M., et al.** (2019). Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of Documentation*, 75(5): 954–76.
- Mustanoja, L.** (2017). Johdanto: kirjeitä sodasta ja rakkautesta. In: Mustanoja, L. (ed.), *Arjen Sirpaleita ja Suuria Tunteita: Kirjeet Sodan Sanoittajina ja Ihmissuhteiden Ylläpitäjänä 1939–1944*. Tampere Studies in Language, Translation and Literature: B5. Tampere, Tampereen Yliopisto, pp. 8–20.
- Nielsen, M. W., Andersen, J. P., Schiebinger, L., and Schneider J. W.** (2017). One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis. *Nature Human Behaviour*, 1: 791–96.
- Nurminen, T.** (2015). “Ehkä se läpimurto on nyt selvä”: sotilaiden kirjeet tapahtumahistorian lähdeaineistona. In: Tikka, M., Taskinen, I. and Nevala-Nurmi, S.-L.(eds), *Kirjeitä Sodasta: Kirjoittamisen Tavot ja Merkitykset Kriisiaikoina*. Postimuseo ja Tampereen Historiallinen Seura, Tampere, pp. 136–51.
- Oberbichler, S., Boros, E., Doucet, A., et al.** (2021). Integrated interdisciplinary workflows for research on historical newspapers: perspectives from humanities scholars, computer scientists, and librarians. *Journal of the Association for Information Science and Technology*, 2021: 1–15.
- Pajunen, M.** (2017). “Sinulla, kultaseni, tuntuu olevan jotakin hampaan kolossa minua vastaan!” Erään avioparin kiistanaiheet sodanaikaisessa kirjeenvaihdossa. In: Mustanoja, L. (ed.), *Arjen Sirpaleita ja Suuria Tunteita: Kirjeet Sodan Sanoittajina ja Ihmissuhteiden Ylläpitäjänä 1939–1944*. Tampere Studies in Language, Translation and Literature: B5. Tampereen Yliopisto, Tampere, pp. 62–93.
- Pirinen, T. A.** (2015). Development and use of computational morphology of Finnish in the open source and open science era: notes on experiences with Omorfi development. *SKY Journal of Linguistics*, 28: 381–93.
- Pirolli, P. and Russell, D. M.** (2011). Introduction to this special issue on sensemaking. *Human-Computer Interaction*, 26(1–2): 1–8.
- Ponkala, M.** (2017). Älä rasita itseäsi liikaa, minulla on täällä ensiluokkainen olo. Kirjeisiin kätketty ikävä nuoren tytön ja äidin kirjeenvaihdossa 1939–1944. In Mustanoja, L. (ed.), *Arjen Sirpaleita ja Suuria Tunteita: Kirjeet Sodan Sanoittajina ja Ihmissuhteiden Ylläpitäjänä 1939–1944*. Tampere Studies in Language, Translation and Literature: B5. Tampereen yliopisto, Tampere, pp. 112–38.

- Riihonen, H.** (2017). “Voi hyvin ja höykyttömästi!” Toveruus, tunteet ja kokemukset miesten välisessä sota-ajan kirjeenvaihdossa. In Mustanoja, L. (ed.), *Arjen Sirpaleita ja Suuria Tunteita: Kirjeet Sodan Sanoittajina ja Ihmissuhteiden Ylläpitäjinä 1939–1944*. Tampere Studies in Language, Translation and Literature: B5. Tampereen Yliopisto, Tampere, pp. 40–61.
- Silén, S.** (2017). Täällä ollessa vasta aukee silmä ja sydän tajuamaan omat tunteensa. Parisuhdediskursseja sota-ajan kirjeissä. In Mustanoja, L. (ed.), *Arjen Sirpaleita ja Suuria Tunteita: Kirjeet Sodan Sanoittajina ja Ihmissuhteiden Ylläpitäjinä 1939–1944*. Tampere Studies in Language, Translation and Literature: B5. Tampereen yliopisto, Tampere, pp. 94–111.
- Small, H. G.** (1978). Cited documents as concept symbols. *Social Studies of Science*, 8: 327–40.
- Taskinen, I.** (2015). Sota-ajan kirjekulttuuri ja kirjeiden kokemusten tulkinta. In Tikka, M., Taskinen, I., and Nevala-Nurmi, S.-L. (eds), *Kirjeitä Sodasta: Kirjoittamisen Tavat ja Merkitykset Kriisiaikoina*. Postimuseo ja Tampereen Historiallinen Seura, Tampere, pp. 12–29.
- Taskinen, I.** (2021). *Social Lives in Letters: Finnish Soldiers’ Epistolary Relationships, Intimate Practices, and Emotionality in World War II*. Doctoral dissertation in History, Yliopistopaino, Joensuu, 416 p.
- Waldvogel, J.** (2007). Greetings and closings in workplace email. *Journal of Computer-Mediated Communication*, 12: 456–77.

Notes

- 1 We use terms female and male in our analysis of gender clues at the lexical level but do not wish to claim that people could be neatly categorized this way. As [Hellinger and Bußmann \(2001\)](#) say, gender is a complex concept, and the terms might be better understood in the light of continuums and variations.
- 2 Omorfi has a large lexicon of lexemes and can analyze also Finnish of the 1940s. It does not perform detailed semantic-level marking of the words, but its repertoire is limited to some semantically oriented tags like ‘male’, ‘female’, ‘title’, ‘currency’, ‘measure’, ‘time’, and ‘country’. Gender-oriented tags male and female were attached to first names of persons only—not to other kinds of words.
- 3 Finnish examples include *laulajatar* (‘a female singer’) and *venakko* (‘a Russian woman’).

- 4 Examples include *lehtimies* (‘journalist’), *työmies* (‘worker’), *esimies* (‘supervisor’). The ending part refers to a male (*mies*—‘man’), but the use of the term is not limited that way.
- 5 Common examples include *Liisa* (female name) and *Lauri* (male name).
- 6 *Vieno* and *Kaino* represent ambiguous names (mainly feminine) used in the 1940s.
- 7 Examples include female names ‘*Ke r t t u*’ and ‘*Helkkyi*’ (*Helkky!*), and male names, such as ‘*O t t o n i !*’ which is a possessive form of *Otto*.
- 8 For example, proper name ‘*Toivo*’ (also meaning hope) was not recognized as a male name.
- 9 For example in signature ‘. . . kisses to Emma and greetings to Esko. Father’ the name *Esko* refers to the male recipient, while father reveals the gender of the sender.
- 10 Examples of Finnish expressions include ‘*äiti*’ (mother), ‘*sisko*’ (sister), and ‘*pikkusisko*’ (compound word for little sister), ‘*isä*’ (father), ‘*setä*’ (uncle), and ‘*veli*’ (brother).
- 11 Examples include ‘*Pikkuäiti*’ (little mother), ‘*mummi*’ or ‘*Mummu*’ (Grandmother), and ‘*isä*’ or ‘*isi*’ (father), and ‘*veli*’ (brother).
- 12 For example, ‘*Väiski*’ may be used as an informal variant of a male name ‘*Väinö*’.
- 13 Examples include suffixes *-si* and *-ni* indicating possession in nicknames ‘*Mappesi*’ (your *Mappe*) (possibly female) and ‘*Ami n i*’ (a noisy form of an expression *my Ami*).
- 14 Examples include ‘*sairaanhoitajar*’ (female nurse) and ‘*laulajatar*’ (female singer).
- 15 For example, ‘*lotta*’ refers to a female member of wartime *Lotta-Svärd* organization.
- 16 Examples of noisy expressions include ‘*Runoi l ijatar*’ referring to (female) poetess, and ‘*vääpel iä*’ referring to (male) ‘sergeant major’. Male gender clues based on military titles included also abbreviations, such as ‘*sot. virk*’ (‘*sotilasvirkailija*’ referring to ‘military functionary’).
- 17 Finnish expressions include variants of word ‘*vaimo*’ (wife) and its dialectical quasi-synonyms such as ‘*eukko*’ (wife/woman); and variants of ‘*mies*’ (husband) with its quasi-synonymic variant ‘*ukkoni*’ and ‘*miehesi*’ (your husband).
- 18 Finnish expression is ‘*vaimoni*’.
- 19 Examples include name variants ‘*Irmeliinus*’ (probably an affectionate form of *Irmeli*) and ‘*Sofirakas*’ (misspelled form of *Sofi* dear), and affectionate variants of ‘kinship terms’, such as ‘*Mamma*’ (variant of mother) and ‘*tyttö*’ (girl) expressing closeness.

- 20 Examples include ‘[rakas] poika’ or ‘isi’ ([dear] boy or daddy) and ‘[rakas] ukko’ referring to ‘[dear] elderly male’.
- 21 Examples include ‘Pappasi’ (your granddaddy) and ‘Mammasi’ (your mom/wife).
- 22 Examples include expressions such as ‘ystävä’ (friend), ‘naapuri’ (neighbour), ‘Lapset’ (children), and ‘pako-laiseti’ (noisy form of refugees!).
- 23 With many regards (‘Monin [terveisin] ...’) contains word ‘Monin’ (With many) which was automatically observed as an inflectional variant of a female name ‘Mona’.
- 24 For example, the recipient was mentioned by name in the farewell: ‘...Eskolle terveiset erikseen. Isä’ (...greetings to Esko individually. Father).
- 25 For example, the noisy expressions ‘Maija l l e’ and ‘Kaa r I na’ obfuscated the female clues *Maijalle* and *Kaarina*, but the relevant documents were found (in both cases). In the first case, the noisy expression accidentally contains the female name ‘Maija’. In the second case the greeting included the idiomatic expression ‘Monin’ (With many) automatically interpreted as a female clue.
- 26 An example of multiple names in greetings include ‘Rakas Emmani ja Eskoni’ (My dear Emma and Esko). An example of the second type entails signature ‘... Näkemiin Eila Meidän kaikkien puolesta Korp. Tauno ...’ (... Goodbye Eila Greetings from all of us Corpr. Tauno ...). The letter was incorrectly retrieved using search criterion FF due to female metadata observed based on female name (Eila) in the signature. Human reasoning in this case allows recognizing the male sender based on the presence of military rank (corpr.) and male name (Tauno) ending the signature.
- 27 Examples include idiomatic expressions With many (‘Monin’), I wish (‘Toivon’), Dear (‘Armas’), and Dearest (‘Kallein’), each of which theoretically could refer to female (Mona) or male proper name (Toivo, Armas, Kalle).
- 28 For example, Finnish acronym ‘Res. kers.’ referring to a male clue ‘reserve sergeant’ appeared in the signature as a noisy string ‘Rea. kers.’ leading to retrieval based on the female name ‘Rea’.