On *ek* and *jak* in Middle Norwegian: mixed methods in historical sociolinguistics

Tam Blaxter, University of Cambridge

Kari Kinn, University of Oslo

1 INTRODUCTION AND BACKGROUND¹

Historical sociolinguistics, like synchronic sociolinguistics, encompasses diverse methods. A robust tradition of quantitative variationist research has developed in the short history of the discipline, represented particularly by work on community and individual variation and change in Early Modern English (e.g. Nevalainen & Raumolin-Brunberg 2003; Raumolin-Brunberg 2009; Evans 2013); even where it focuses on individual usage, this approach typically works with variation between texts. This is matched by qualitative research, especially as we approach the border of historical (socio)pragmatics (e.g. Mazzon 2003; Culpeper 2009; Hansen 2009); such work pays more attention to variation within texts. In this paper, we will use a case study to argue for uniting these approaches. By investigating the same variable phenomena from both intra- and inter-textual standpoints and with both quantitative and qualitative methods, we can reach a richer understanding of historical language use in its social context than with either approach alone.

Our case study is ‘breaking’ in the 1st person sg. nom. pronoun in the history of Norwegian. In the present-day Mainland Scandinavian languages, the 1st person sg. nom. pronoun takes different forms. Swedish and Danish use the ‘broken’ (i.e. diphthongised) forms <jag> and <jeg>; in spoken and written Norwegian, on the other hand, there is variation between broken and unbroken forms. Of the two official Norwegian written standards, Bokmål has broken <jeg>, while Nynorsk uses unbroken <eg>; cf. figure 1 for an overview of the spoken forms and their geographical distribution.

The traditional Old Norwegian form of the 1st sg. pronoun is *ek*, i.e. an unbroken form (Noreen 1970:309); this sets Old Norwegian apart from Old Swedish and Old Danish, which had broken forms such as *jæk/jak* already in the earliest records (Noreen 1904:388). The introduction of broken forms into Norwegian has been ascribed to influence from Danish and Swedish (see e.g. Hødnebo 1971, Indrebø 2001:193, Wetås 2008:234 and Berg 2013:64ff).² However, we are not aware of any

1 Our warm thanks to the audience at the GLAC 2016, Tor Weidling at the National Archives of Norway, Oslo, Ivar Berg at the Norwegian University of Science and Technology, Trondheim, and to Karl Gunnar Johansson at the University of Oslo for their helpful comments on this research.

2 Seip (1954b:220ff), on the other hand, takes the view that broken forms of the 1st sg. nom. pronoun have existed in East Norwegian dialects for as long as they have in Danish/Swedish, although he also suggests that the increased use

large-scale, empirical investigations of the change in Norwegian, and with the present study we also aim to fill this gap. We will investigate how broken forms of the 1st person sg. nom. pronoun arose in Norwegian and the distribution of broken and unbroken forms in medieval Norway, taking into account geographical, social and textual factors.

The study is based on charters from the *Diplomatarium Norvegicum* (DN), a body of charters from (or about) medieval Norway. In all, the DN contains around 20,000 documents of which 13,056 are (at least partially) in a Nordic language. In the present study, however, we only include the charters that are originals, in all 10,683 documents containing 2,761,765 words of text.

The duality of approach in historical sociolinguistics finds its echo in the Norwegian philological tradition. Many studies in the history of Norwegian are qualitative: they examine a small number of documents in great detail, aiming to trace and explain variation and changes in language use at the levels of speakers, scribes and localities (e.g. Hægstad 1902, Knudsen 1927, Vågslid 1969, Rindal 1979). There is also a real tradition of large-scale, quantitative research on this material, as studies like Mørck (1980) and Berge (1974) attest; such studies describe differences between regions, social classes and periods that can only be seen in larger datasets. Both approaches have much to offer. While qualitative work is needed to fully describe the fascinating and ephemeral phenomena in individual language use from which societal patterns emerge, quantitative work is needed to tell the sweeping stories from which language histories are compiled. In this study we aim to unite these approaches, using the findings of each to inform the other. In doing so, we will identify valuable new insights into contact and change in medieval Scandinavia and thus demonstrate the value of our combined methodology.

The paper is organised as follows: in section 2 we discuss the introduction of broken forms on a macro level, looking at large quantities of data and using quantitative methods. In section 3 we discuss the distribution from a micro-level perspective, looking closely at individual documents and signatories. In section 4 we discuss our findings and conclude the paper.

of this feature in documents from the middle of the 14th century onwards might to some extent be triggered by Swedish influence.

Legend

Form of the first person singular subject pronoun:

- breaking (je, jæ, jæi, ja, jag)
- ei, æi(g)
- i
- ig, ik, i
- æ(g), e(g)

Figure 1: Forms of the 1.sg. subject pronoun in the modern North Germanic languages, reproduced from Sandøy (1985)

2 QUANTITATIVE APPROACH

2.1 Introduction

Quantitative investigation of variation in a large dataset has several advantages. Using inferential statistics we can first identify which external factors exert significant influence on the choice of variant. This allows us to avoid over-interpreting the data or attempting to explain apparent effects which are, in fact, simply the result of noise, an important procedure when dealing with the typical 'bad data' of historical linguistics. We can use descriptive statistics to investigate the exact form of those patterns identified and posit explanations for these population-level patterns. These explanations can then both inform and be tested by our micro-scale, qualitative investigations.

2.2 Chronology

All instances of 1st sg. nom. pronouns were identified and classified according to whether they showed breaking (e.g. <ek> vs. <iak>) and voicing (e.g. <iak> vs. <iegh>)³ ⁴. Figure 2 shows the rates of these different orthographies in the corpus by 50-year moving window. As can be seen, at the beginning of the period characteristically Old Norwegian forms predominate, with no breaking of the vowel or voicing of the consonant. Breaking spreads rapidly from the mid-14th century onwards, while voicing of the final consonant spreads rapidly from the latter part of the 15th century.

3 Spelling conventions in Old and especially Middle Norwegian are complex and inconsistent, creating problems of interpretation for many phenomena. However, for the purposes of breaking, the situation is relatively straightforward: broken forms are those which begin with an explicit indication of a glide (that is one of <y>, <j> or <i>); unbroken forms begin directly with the vowel (one of <a>, <e> or <æ>).

4 Breaking and voicing in the 1st sg. nom. pronoun may be connected in that both may be due to East North Germanic influence, but are fundamentally different variables. This paper will concentrate only on breaking.

Figure 2: Proportions of orthographies for the 1.sg.nom. pronoun in the DN by 50-year moving window

2.3 Independent variables

Each charter has been tagged for a series of pieces of metadata which can be used as predictors.

2.3.1 Date

Each charter has been tagged for its date of publication. In the case of most charters, this date is given at the beginning of the body or in the eschatocol. Usually, charters express the year in terms of the number of years the current monarch has reigned; less often, years are given since the birth of Jesus Christ. Days are typically expressed relative to ecclesiastical festivals. Two examples are given below.⁵

- (1) [...] þetta bref er giort var j Biorghvin Margrettar messv aftan aa avdru ok tutughta rikiss 'okkars vyrdhvilgix herra Hakonar med gvds nad Noregx konvngs.
'[...] this charter which was made in Bergen on [the] eve of [the] mass of [St.] Margaret in [the] 22nd [year] of [the] reign of our worthy lord Hakon with God's grace king of Norway.'
(Eschatocol from DN I.445, 19th of July 1377)
- (2) Scriffuet j Spioteberg sancti Jacobi dag anno domini mcdnonagesimo.

⁵ We standardise place names, but generally not personal names, in the English translations.

‘Written in Spydeberg on St. Jacob's day AD 14 ninety.’
(Eschatocol from DN III.973, 25th of July 1490)

The published editions of the DN provide Gregorian calendar interpretations of these dates for each text and these were used for this study. Dates were encoded as a number representing the year. Where a range of possible dates was given for a text, the middle of this range was used.

2.3.2 *Social rank*

The vast majority of charters in the DN begins with formulae which introduce the signatories by name and titles. An example is given below.

- (3) [...] At jak andres benktsson kaniker j hambre och soknaprester j Ringabo amunder grimsson oc torsten pædhersson suornir mæn oc mykid annad got folk: [...] ‘[...] That I, Andres Benktsson, canon in Hamar and parish priest in Ringeby, Amunder Grimsson and Torsten Pædhersson, sworn men, and many other good people [...]’ (Protocol from DN XXI.497, 1458)

Each charter was tagged for the titles of its first signatory⁶ and these titles were grouped into ranks. The classification of titles into ranks is identical to the one suggested by Mørck (1980:16–17; 1999:272–274) with the sole modification that Mørck's category of 'bønder' ('farmers') was separated into 'magistrates' (people with the title *lögrettismaðr*) and 'untitled individuals'. This approach was taken on the assumption that, contra Mørck, documents whose signatories are not identified with any title are not, in fact, a unitary social category in the DN (documents by individuals with no title) but a socially heterogeneous group of charters in which the signatory's title happens not to have been given⁷; by contrast, charters by *lögrettismenn* are assumed to represent a meaningful social category. The social categories used and the number of documents in the sample thus categorised were:

1. royalty (3 charters)
2. higher clergy (46 charters)
3. higher officials and the aristocracy (368 charters)
4. lower officials (185 charters)
5. citizens and town officials (30 charters)
6. lawyers (263 charters)
7. lower clergy (321 charters)
8. magistrates (300 charters)
9. untitled individuals (1254 charters)

⁶ This rests on the assumption that the first signatory is the best candidate for the person we should consider the 'speaker' in these documents: they are the identifiable individual most likely to have had an influence over the form of the language in the charter. This is rejected by some previous scholars (e.g. Pettersen 1975:66; Tønnessen 1995:12–13) on the basis that unknown scribes might have been employed to write documents on behalf of known signatories, but even in such cases signatories may have influenced textual language through dictation (Mørck 1999:271–272) and the substantial success in previous studies in identifying linguistic correlates of first signatories (Mørck 1980; Farstad 1991; Mørck 1999:199; Blaxter 2017) is cause for optimism.

⁷ For more detail, see Blaxter (2017).

10. problem category (30)

11. unclassified (13)

Note that the low cell counts for the royalty and higher clergy categories is largely a result of the fact that writers in these social groups usually used the royal plural and thus rarely provide tokens of the first person singular pronoun. The 'problem category' contains documents whose first signatories are hard to classify, largely made up of titles which occur only once in the data. The 'unclassified' category contains documents with no signatories.

2.3.3 Name

Each charter was tagged with the name(s) of the first signatory and each of these names was categorised according to its etymology. The etymological classes were: North Germanic, Biblical, non-Biblical Greek/Latinate, German and unclassified/other. The charters were then tagged with these categories or as 'mixed' if the first signatory had multiple names of different etymological sources. The categories and counts were as follows:

1. German (54 charters)
2. Greek/Latinate (16 charters)
3. Biblical (43 charters)
4. Mixed (1066 charters)
5. Nordic (1556 charters)
6. unknown/unclassifiable (78 charters)

The assumption behind this tagging scheme is that, in some instances, the writer's name will indicate their ethnicity (as German or Scandinavian) and may correlate with their social background (with higher status individuals more likely to have German, Biblical or Greek/Latinate names). Note that it is *not* assumed that these associations are categorical. Clearly, especially in the Late Medieval and Early Modern periods, German names began to be used by native Norwegians and some Biblical and classical names had relatively wide currency. Nevertheless, if these correlations are present at some significant level in the data, they may be used to illuminate the effects of ethnicity and social class.

2.3.4 Gender

Each charter was tagged on the basis of signatory names both for whether the first signatory was female (101/2813 documents) and for whether at least one signatory was female (165/2813 documents).

2.3.5 Localisation

A subset of documents in the corpus were localised to one or more specific localities. Sources of geographical information in charters include:

1. place of publication in the eschatocol, such as:

(4) [...] þetta bref er gort var j Hamre [...]
'[...] this charter which was made in Hamar [...]'

(Eschatocol from DN II.510, 4th of April 1388)

2. place at which the legal meeting took place stated at the beginning of the body, such as:

- (5) [...] varom vit j hia Jon ok Hælgj Simunar synir a inra Hundæidi j Nordfiordom, er [...] ' [...] we were there, Jon and Hælgj Simunaron, at Hundeidet indre in Nordfjord, when [...]'
(DN III.122, 28th of April 1321)

3. place of origin of the signatories given in the protocol, such as:

- (6) Þæt se ollum godum monnum kunnit at ek Haraldær Jonsson logman a Vplanddum [...] 'Be it known to all good people that I, Haraldær Jonsson, lawyer in Oppland [...]'
(Protocol from DN III.588, 31st of July 1408)

4. location of land being sold or of other events described in the charter.

The editors of the DN provide limited localisation information (typically a single-word place name) for each charter. Although no explicit account is given, this localisation appears to be based on these sources of information in this order of preference. Thus the same procedure was followed here: sources of information 1. to 3. were each recorded where present, and source 4. only when none of the other sources were available. Where a charter was localised to multiple places, it is mapped in this paper as the centre point of those localisations. For more detail on these processes see Blaxter (2017). 2010 out of 2814 documents used in this study were localised.

2.4 Non-geographical regression

A binomial regression model of the likelihood of finding broken forms was fitted using the `glm` (generalised linear model) function from the `lme4` package in R (Bates et al. 2016). The dependent variable was the proportion of broken forms in each document, weighted by the number of tokens. In addition to the independent variables described above, the model included two pieces of geographical information: whether at least one of the localisations of a charter was urban (counting the five largest towns, Oslo, Tønsberg, Trondheim, Stavanger and Bergen, as urban localities) and whether at least one of the localisations was in Sweden or one of the two regions most closely connected with Sweden (Bohuslän and Jämtland). The former variable was included as an initial test for hierarchical diffusion: if breaking was spreading by hierarchical diffusion (as might be predicted for a variable spreading partly as a purely written phenomenon from an external variety spoken by a wealthier and more powerful population), it might be expected to be found first in cities where contact was more intensive. The latter was included as an initial test of the prediction that breaking spread from Swedish: if so, it would be expected to be found earlier in charters with some connection to Sweden.

All of the independent variables were found to have a significant effect on the dependent variable with the exception of `urban_connection`; as the `drop1` function showed that including this did not significantly improve the predictive power of the model, it was excluded. The resulting model is displayed as Figure 3.

```

Call:
glm(formula = prop_breaking ~ mid_year + first_sender_rank +
    first_sender_name_type + at_least_one_sender_is_female +
    swedish_connection_b, family = "binomial", data = ek_data[which(!ek_data$probably.exclude),
    ], weights = sample_size)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.7730  -0.5603   0.1172   0.5832   8.3866

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -69.912305    1.459988 -47.886 < 2e-16 ***
mid_year         0.048506    0.001009  48.090 < 2e-16 ***
first_sender_rank=citizen    0.075822    0.329588   0.230 0.818053
first_sender_rank=higher_official 0.688433    0.137583   5.004 5.62e-07 ***
first_sender_rank=higher_clergy  0.608464    0.393932   1.545 0.122445
first_sender_rank=royal      7.183784    1.194965   6.012 1.84e-09 ***
first_sender_rank=lower_official -0.334671    0.136829  -2.446 0.014449 *
first_sender_rank=lower_clergy  -0.352901    0.119225  -2.960 0.003077 **
first_sender_rank=lawyer     -0.430797    0.141213  -3.051 0.002283 **
first_sender_rank=magistrate  -0.327597    0.109694  -2.986 0.002822 **
first_sender_rank=problem_category 2.066528    0.572011   3.613 0.000303 ***
first_sender_rank=unclassified -0.575663    0.856027  -0.672 0.501277
first_sender_name_type=biblical  0.594156    0.318766   1.864 0.062332 .
first_sender_name_type=german    1.810433    0.343525   5.270 1.36e-07 ***
first_sender_name_type=latinate  1.547953    0.989941   1.564 0.117892
first_sender_name_type=nordic   -0.107020    0.074690  -1.433 0.151901
first_sender_name_type=unclassified -0.564193    0.345830  -1.631 0.102803
at_least_one_sender_is_female  0.572285    0.171388   3.339 0.000840 ***
swedish_connection    3.430357    0.183915  18.652 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 15011.3 on 2787 degrees of freedom
Residual deviance: 4754.5 on 2769 degrees of freedom
AIC: 4994.5

Number of Fisher Scoring iterations: 6

```

Figure 3: Output of non-geographical regression

The regression shows, just as we would expect, that a later text date (`mid_year`) favours breaking: there is increase in the broken forms over time during the period covered by the corpus. In terms of social groups, we see that documents by royalty, higher officials and nobles and the 'problem category' all favour breaking compared with documents by untitled individuals (the reference level for the factor), whereas magistrates, lower officials, lower clergy and lawyers all disfavour it. Thus we seem to have a neat division of the ranks into the higher classes, who favour the change, and the lower classes, who disfavour it. The evidence from names is more equivocal: writers with German names strongly favour the change compared with writers with mixed names (the reference level for the factor); writers with Biblical names also favour it, although this effect does not quite reach significance. If we change the reference level to 'Nordic', we find the following coefficients:

<code>first_sender_name_type=mixed</code>	0.107020	0.074690	1.433	0.151901
<code>first_sender_name_type=biblical</code>	0.701176	0.317367	2.209	0.027150 *
<code>first_sender_name_type=german</code>	1.917453	0.341665	5.612	2.00e-08 ***
<code>first_sender_name_type=latinate</code>	1.654973	0.989355	1.673	0.094371 .
<code>first_sender_name_type=unclassified</code>	-0.457173	0.345337	-1.324	0.185554

Figure 4: Coefficients for releveled 'name' factor

Thus we can see that writers with Biblical or German names favour breaking compared with writers with Nordic names.

Gender also has a significant effect, with documents with at least one female signatory favouring breaking compared to those with none. Finally, documents with at least one localisation in Sweden, Bohuslän or Jämtland strongly favour the change compared with those localised only within Norway proper.

2.5 Geographically weighted regression

A geographically weighted regression (GWR) was fitted using the GWR4 software (Nakaya et al. 2014). GWR has previously been used for historical dialectology by Willis (2017). As the software does not have as sophisticated array of options as the `glm` function in R, the dependent variable was boolean: whether there was at least one token with breaking in each document. The date, first signatory rank, the first signatory's gender and the first signatory's name were included as independent variables. As GWR4 only allows for numerical independent variables, first signatory rank was coded as a scale of increasing ranks from 0 to 10: magistrates, unclassified, lower clergy, lower officials, lawyers, untitled individuals, citizens, higher clergy, higher officials and nobles, problem category, royalty. This was judged reasonable as there is independent reason to conceive of the social classes as an ordered list. By contrast, the name categories in no sense constitute an ordered list, and thus it would not be a valid approach to treat them each as numerical levels; accordingly, this variable was collapsed into Nordic names (1) vs. all others (0).

In Willis' study, two assumptions were built into the model: that the slope of the change was the same at all localities, and only the intercept (i.e. the year the change began) varied geographically; that the effects of other predictors did not vary between localities. Thus the only local coefficient (that is, coefficient allowed to vary across localities) was the intercept, and all other coefficients were global. In this study, it was assumed that the slope of the change was the same at all localities, thus the date was included in the model as a global coefficient. However, diverging from Willis' approach, all other independent variables were first tested with local coefficients. An initial run of the model indicated that only the intercept and first signatory rank showed statistically significant geospatial variation (that is, only for these two variables did the local model show a significant improvement in predictive power compared with the global model). Thus in the final model, only these two were treated as local coefficients and the other two variables were treated as global coefficients. The output of this model is shown in Figure 5.

```

*****
GWR (Geographically weighted regression) result
*****
Bandwidth and geographic ranges
Bandwidth size: 260.328289
Coordinate      Min      Max      Range
-----
X-coord      -19.358425      18.068581      1987.580929
Y-coord       52.715385      69.362161      1851.039682
(Note: Ranges are shown in km.)

Diagnostic information
Effective number of parameters (model: trace(S)): 12.248550
Effective number of parameters (variance: trace(S'WSW^-1)): 10.683370
Degree of freedom (model: n - trace(S)): 1987.751450
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)): 1986.186270
Deviance: 1035.907427
Classic AIC: 1060.404528
AICc: 1060.567885
BIC/MDL: 1129.007463
Percent deviance explained 0.625587

*****
<< Fixed (Global) coefficients >>
*****
Variable      Estimate      Standard Error      z(Estimate/SE)
-----
year_from_1442      3.477861      0.167345      20.782623
first_sender_is_female      0.192291      0.089155      2.156821
native_name      -0.229573      0.091018      -2.522290

*****
<< Geographically varying (Local) coefficients >>
*****
Summary statistics for varying (Local) coefficients
Variable      Mean      STD
-----
Intercept      0.112159      0.418233
rank_number      0.311056      0.212842

Variable      Min      Max      Range
-----
Intercept      -0.472476      2.076470      2.548946
rank_number      0.013034      1.085103      1.072069

*****
GWR Analysis of Deviance Table
*****
Source      Deviance      DOF      Deviance/DOF
-----
Global model      1143.958      1995.000      0.573
GWR model      1035.907      1986.186      0.522
Difference      108.051      8.814      12.259

*****
Geographical variability tests of local coefficients
*****
Variable      Diff of deviance      Diff of DOF      DIFF of Criterion
-----
Intercept      92.927621      3.511364      -85.904894
rank_number      16.915323      3.579619      -9.756085

```

Figure 5: Output of geographically weighted regression

As can be seen from this from this figure, a later date, a female signatory and higher social ranks favour broken forms; this matches the findings of the non-geographical regression. A signatory with a Nordic name favours unbroken forms.

The two local coefficients, for the intercept and for first signatory rank, are mapped in Figure 6.

Figure 6: GWR local coefficients

As can be seen from the left map, the change to broken forms is found earliest in Jämtland and Trøndelag, later in eastern Norway and latest in the west. From the right map we can see that the influence of rank is strongest in those places that the change reaches first as well as Oslo and Akershus, and weakest along the west coast. The degree of variation is relatively large: at its weakest, on the west coast around Bergen, signatory rank has no effect on breaking at all, whereas in Oslo, Trøndelag and Jämtland it is a relatively strong predictor.

2.5.1 Plots of independent variables

Figure 7, Figure 8 and Figure 9 show the breakdown of the change by time period and gender, first signatory name and first signatory rank respectively. Figure 10 shows a kernel density estimation with a Gaussian kernel in time with bandwidth (standard deviation) of 60 years and an adaptive uniform kernel in space with a bandwidth of the weighted equivalent of 50 nearest neighbours⁸.

Figure 7: Proportion of broken forms of the 1.sg.nom. pronoun in the DN by 50-year moving window and gender of the first signatory

⁸ For an example of the use of KDE in (synchronic) dialectology, see Rumpf et al. (2009); for more detail on this method, see Blaxter (2017).

Figure 8: Proportion of broken forms of the 1.sg.nom. pronoun in the DN by 50-year moving window and first signatory name

Figure 9: Proportion of broken forms of the 1.sg.nom. pronoun in the DN by 50-year moving window and first signatory social group

Figure 10: KDE for breaking in the 1.sg.nom. pronoun 1350-1500

As indicated by the regression analyses, Figure 7 clearly demonstrates that the change to broken forms was led by documents signed by women. This may indicate that the change in the spoken language which took place in the east was led by women. Similarly Figure 8 confirms that the change was largely led by writers with German, Latinate and Biblical names. Figure 9 slightly complicates the finding that the change was led by writers of higher rank: some higher groups (higher officials and nobles, higher clergy) clearly lead the change and the middle classes (lawyers, lower officials) lag behind. The lowest social classes (magistrates, lower clergy) take longest to complete the change but are relatively innovative in its earliest years. Surprisingly, the untitled individuals group, primarily representing documents whose signatories' social backgrounds are unknown, is consistently innovative.

Figure 10 confirms the findings of the geographically weighted regression presented in Figure 6, viz., that the change took place earliest in Jämtland, followed by Trøndelag, Bohuslän, Hedmark and other parts of eastern Norway. The west coast was the last region to be reached.

2.6 Discussion

The results presented above paint a very clear picture: we see a change spreading from east to west and from north to south, led in the eastern part of Norway by the higher social classes and by female writers, with less social differentiation in the west. It is interesting to note that this differential social differentiation separates eastern Norway, where the change is known to have eventually taken place in the spoken language, from western Norway, where it did not. Thus we might conclude that the change in the spoken language was socially structured, as we expect from modern sociolinguistic studies of ongoing changes, distinguishing it from the purely written change in the west.

These results support the supposition that the broken forms in Norwegian stem from contact with Swedish. Although Jämtland and Bohuslän might be expected to show greater influence from written and spoken Swedish and thus their innovative statuses are not conclusive evidence of the source of the change in Norwegian, the fact that Trøndelag, Østfold, Hedmark, Vestfold, Akershus and Oslo, the places with the closest links with Sweden, are also leading areas is strongly suggestive of a role for contact. Similarly the fact that writers with German names can be seen to have led the change and those with Nordic names to have lagged behind seems to evidence a contact-driven account.

From a methodological standpoint, these findings clearly speak to the value of the quantitative approach. It would be impossible to arrive at such confident generalisations about language use from impressionistic description, given the large number of texts distributed so unevenly over such a large period and area. Even if broad patterns such as the innovative status of eastern texts or texts with signatories from the higher social classes could be identified, it would be impossible to disentangle and quantify these effects using qualitative methods. Only with a quantitative, inter-textual approach could we have arrived at these findings.

3 QUALITATIVE APPROACH

In section 2 we took a quantitative, macro-level approach to the introduction of broken forms of the 1st sg. nom. pronoun. In this section we will exploit the possibilities offered by our annotated corpus to study variation on a micro level. In what follows we will explore variation between broken and

unbroken forms from a qualitative perspective, focussing first on variation within individual charters (section 3.1), and then on individual signatories (section 3.2). In section 3.3 we provide a brief summary.

3.1 *Intra-textual variation*

Typically, variation between broken and unbroken pronoun forms of the 1st sg. nom. pronoun is found on the inter-textual level. However, there is also a certain amount of intra-textual variation, i.e. broken and unbroken forms coexisting within individual charters. Previously, this has been observed by Hødnebo (1971), but we are not aware of any more extensive studies of the phenomenon.

We have found intra-textual variation between broken and unbroken pronoun forms in 97 charters, and in what follows, we will look at some of them in more detail. Our aim is to identify factors that might have contributed to intra-textual variation and present cases in which each of the factors can be relatively clearly discerned. The factors to be discussed are formulaic language, direct or reported speech, point of view and, finally, co-authorship or time gap. Note that we are not attempting to account for every instance of intra-textual variation; in many of the relevant charters more than one explanation is possible, or multiple factors seem to be interacting, sometimes in complex ways. We have also found variation that does not neatly fall into any of the categories that we are proposing.⁹

3.1.1 *Formulaic contexts*

As is well-known, most charters include formulae that are to some extent fixed. Formulaic language is particularly, but not exclusively, associated with the protocol and the eschatocol (see Hamre 1972). Occasionally, the protocol, eschatocol or other formulaic contexts differ from (the rest of) the body text in terms of broken vs. unbroken pronoun forms. This kind of variation might stem from conventions that were not productive apart from in formulae; one way in which such conventions could be promoted is through the use of exemplars.

Interestingly, formulaic contexts as such do not clearly favour either pronoun form; we have found documents with broken forms in formulaic contexts and unbroken forms elsewhere, but also, conversely, unbroken forms in formulaic contexts and broken forms elsewhere. Moreover, there are cases in which formulaic language and exemplars might have had an influence on the pronoun forms, but in a rather intricate way; we will come back to this towards the end of this subsection.

A charter that exhibits a relatively clear pattern is DN XII.210 (1446). This document displays unbroken pronoun forms in the introductory formula and in the eschatocol (altogether 2 instances), but broken forms in the body text (2 instances); the variation is illustrated in Extract 7.¹⁰

- (7) a. Thes kennest **ek** Jon Niclisson laghman ij Throndem med thesso mino opnno brefe at **jak** war a Frosto thingstad rettom, laughardagen efter sancte Botulfs dagh [...]
'I, Jon Niclisson, lawyer in Trondheim, acknowledge with this open letter that I was at Frosta, the right place of assembly, on the Saturday after St. Botolph's day [...]'

⁹ Cf. for instance DN II.817, DN III.839 and DN I.774.

¹⁰In this and the following extracts 1st sg. nom. pronouns appear in boldface; emphasis is ours.

- b. sagde **jak** theim sex manna dom [...]
 - ‘I told them the verdict of six men [...]’
- c. Til mere visso her vm sæter **ek** mitt jnsighle med thessom fornemdom mannom, fore thetta bref [...]
 - ‘For additional proof I put my seal on this letter, along with the aforementioned men [...]’
 - (DN XII.210, 1446)

The passage rendered in 7a is the beginning of the document. The charter does not have a protocol, but starts directly with the *dispositio* (see Hamre 1972:50ff); nevertheless, the construction <Thes kennest ek [...]> ‘I acknowledge that [...]’ is formulaic in the sense that it occurs in numerous charters as an introduction to different kinds of legal statements. It is worth noting that the distance between the unbroken form <ek> in the introduction and the first broken form <jak> is very short; the switch to <jak> is intra-sentential and seems to happen as soon as the scribe turns to the specifics of the case in question.

The charter is localised to the area around Trondheim (the signatory is a lawyer in <Thromdem>, and the document deals with land in Skamfer (<Skaumfarar>) [in Rennebu, Sør-Trøndelag]). At the relevant time (1446), charters from this area had a relatively high proportion of breaking (cf. Figure 10). It thus seems likely that the unbroken forms in the introduction and eschatocol are archaic; they reflect older writing conventions and are possibly preserved through exemplars.

In DN II.947 (1488), we find a pattern which is in some respects the opposite of the one just described. In this charter, the broken form <Jeg> is used once, in the introduction of the charter, while the unbroken form <ek> is used elsewhere, both in the body text and the eschatocol, altogether 5 times. The variation is illustrated in Extract 8.

- (8) a. **Jeg** Eileff Olafson prester a Spioteberge och profaster i ofrelutanom i Borgæsyslo kungiør med tessu mine opne brefwe [...]
 - ‘I, Eileff Olafson, priest in Spydeberg and provost in upper Borgesyssel announce with this open letter [...]’
- b. var **ek** och j Sarfsborg vm Pædersmesso
 - ‘I was also in Sarpsborg at St. Peter’s mass.’
- c. Thil mere sannind oc visssen tha trycker **ek** mit incigle fore thette [...]
 - ‘For additional proof, I put my seal on this [letter] [...]’
 - (DN II.947, 1488)

The passage cited in 8a is the beginning of the charter; this is another document that starts directly with the *dispositio*. However, in contrast to in the document from the Trondheim area discussed above, the pronoun form in the introduction does not stand out as more conservative than the rest of the charter – it is the older form <ek> that seems to be productive. The signatory is a priest in Spydeberg and provost in upper Borgesyssel [Østfold]; in this area, the proportion of broken pronoun forms is generally lower than in the surrounding areas at the relevant time (cf. figure 10). Although the general picture is not quite clear, it is worth noting that charters from this area and period display certain other conservative features: the loss of <þ> took longer here than in

surrounding regions (Blaxter 2017:236-237), as did the loss of the 1.pl. *-um* on verbs preceding their subjects (Blaxter 2017:302). On this backdrop, <ek> as the predominant form could possibly be seen as a part of a locally preserved norm. The broken form in the introduction seems to follow a more specialised convention, but the source of this convention is probably not the traditional forms of the area; perhaps it was inspired by impulses from the adjacent cities.

Another charter that has a broken form in the introduction, but unbroken forms elsewhere, is DN X.271 (1490). The broken form is spelt <jeg>; the unbroken forms are spelt <eg> and occur three times, twice in the body text and once in the eschatocol. The variation is illustrated in Extract 9.

- (9) a. Teth kœnnas **jeg** Torræ Redarsson med tethæ mit oppænæ bref at **eg** hauer solth Suænungh Rolfson jfff markæbol jordh j Lonir [...]
 ‘I, Torræ Redarsson, acknowledge with this open letter that I have sold 4 *merkrból* land in Lone to Suænungh Rolfson [...]
- b. kænæs **eg** firnæn^d Toræ hauæ opbore(t) først pænig oc øvæst (oc) allum te emelem er j kop vort kom
 ‘I, the aforementioned Toræ, acknowledge that I have received the full payment that was a part of our deal.’
- c. til ytermer vissa bedes **eg** dandæ men jnseylæ firæ tethe bref [...]
 ‘For additional proof, I ask for the seals of the good men [...] on this letter [...]
- (DN X.271, 1490)

Once again, we see that the distance between the broken form in the introduction and the first unbroken form in the body text is very short (cf. 9a). This corroborates the idea of special conventions applying to the introductory formula. DN X.271 is located to Telemark; it deals with land in a farm in Tørdal (<Tørædal>). In charters from this area, breaking was rather common at the relevant time (cf. Figure 10), although it does not seem to be productive in DN X.271. A plausible explanation for the predominance of unbroken forms in this charter is influence from the spoken language. Tørdal is located in a dialect area which preserved the unbroken form of the 1st sg. nom. pronoun until modern times.¹¹ Moreover, the spelling of the pronoun is <eg>, and not the traditional Old Norwegian form <ek>; this can be taken to suggest that the form is based on the local dialect. Finally, it is worth noting that the charter is characterised by multiple omissions and amendments. This might indicate that the scribe had a low degree of literacy or was in a hurry; both of these factors might have made paved the way for influence of the spoken dialect.

The final case to be discussed here is DN XII.143 (1409), which exhibits an intriguing mix of forms. The unbroken form <ek> is used in the introduction, whereas the broken form <jak> is used twice in the body text; cf. Extract 10.

- (10) a. Thet se alle ollom godhom monnom kunnekt them sem thetæ bref see ædher høeræ kienest **ek** Jans Olaffzsson at **jak** hæffuir selt Bottolf Aluersson at Ærtzgardom alt Bue [...]
 ‘It shall be known to all good people who see or hear this letter that I, Jans Olaffzsson, acknowledge that I have sold all of Bu to Bottolf Aluersson at Ersgård [...]

11 This is not easily discerned from the map in figure 1; however, the Nordic Dialect Corpus (Johannessen et al. 2009) provides a more detailed picture. Tørdal is not a recording point itself, but it is located between Drangedal and Nissedal, two places in which the 1st sg. nom. pronoun form is unbroken (transcribed as *e*, *eg* or *æ*).

- b. Thet bekennisz **jak** hæffuir wpburit minste pening ok mestom som j okkart kiøp kom
'I acknowledge that I have received the full payment that was a part of our deal.'
(DN XII.143, 1409)

At first glance, this looks similar to the charter discussed in the beginning of this subsection, DN XII.210, for which we argued that the unbroken forms were a conservative feature retained in certain formulae. In the present case, it may very well be correct that the form <ek> in the introduction has support in exemplars or particular conventions; however, there are some complicating factors. First, the charter is located to Fåberg [Oppland], an area which had a low proportion of breaking at the time when the charter was issued (cf. figure 10); thus, we do not immediately expect the broken form to be productive and the unbroken form to be particularly conservative. Second, there are some indications that the scribe's level of literacy was not very high: there are two instances of what look like erroneously copied formulae,¹² and a spelling error (<monom> for *minom*). In addition to the broken pronoun forms, the charter displays other signs of Swedish lexical influence (e.g. the prefix *be-* in <bekennisz>). It is possible that the distribution of <ek> and <jak> in DN XII.143 reflects a more or less incidental mix of different sources of influence on the written language of an inexperienced scribe, rather than a genuine opposition between productive and more formulaic forms.

3.1.2 Direct or reported speech

In some charters, variation between broken and unbroken pronoun forms correlates with direct or reported speech. The forms used in direct or reported speech might reflect the spoken language of the quoted person, which may differ from the language used in other parts of the charter.¹³ We have found two subtypes of variation associated with direct or reported speech: first, there are cases in which the 1st sg. nom. pronoun form in direct/reported speech differs from the pronouns found elsewhere in the document. Second, there are cases in which the pronoun forms in the speech of different persons vary.

The first type of variation is instantiated in the charter DN I.740 (1432, Mo [Telemark]). In this document, <ek> is the most common pronoun form; it is used both in the body text and in the eschatocol, altogether six times, while <iek> is used twice. Both occurrences of <iek> are found in a passage quoting a man called Towe Jonsson, who has been killed.¹⁴ The passage is rendered in Extract 11.

12 Cf. Extract (10a), where the formula starting with <Thet se alle ollom godhom monnom kunnekt them sem thetæ bref see ædher hæræ [...]> 'It shall be known to all good people who see or hear this letter [...]' appears to be unfinished. This introduction is normally followed by an embedded clause introduced by *at* 'that', specifying the contents of what shall be known, not a new main clause, as in this case. Somewhat similarly, we would normally expect an *at*-clause, not a new main clause, to follow <Thet bekennisz **jak**> in 10b.

13 Hødnebo (1971:149–150), in a discussion of charters as linguistic evidence, mentions a case in which variation between *ek* and *jak* in the direct speech of different persons is unlikely to be due to differences in the spoken language (the relevant charter is DN I.699, 1424). However, this does not mean that the idea of quotations reflecting speech should be discarded altogether.

14 As mentioned in footnote 3, Old and Middle Norwegian spelling can be rather inconsistent. When discussing named persons, we generally base our spelling on the first occurrence of that name in the relevant charter.

- (11) [...] en Towe fyrnemder swaradhe, huat hauer **iek** lofuet thic, thet **iek** hauer ey haldet
‘[...] and the aforementioned Towe replied: what have I promised you that I have not kept?’
(DN I.740, 1432)

In the context of the findings presented in section 2, it would be interesting to know if Towe’s social background differs from that of the signatory, who is a *lénsmaðr* (‘sheriff’), i.e. a lower official. A second possibility is that Towe and the signatory, Saudulfuer Þordhasson, are from different places: it might have been that Towe was from northern or eastern Telemark, which had higher rates of breaking in this period, and Saudulfuer from further west. Since Saudulfuer is not known from any other documents and Towe is mentioned just once elsewhere, in the charter DN XV.68 which records the resolution of the case begun here, we have no additional information to disambiguate these possibilities.

The second type of variation, i.e. variation in the speech of different persons, is found in DN I.760 (1437, Tolstad [Oppland]). This charter contains 3 instances of the unbroken form <ek> and 1 instance of the broken form <jek>. All the instances of <ek>, two of which are rendered in Extract 12, are found in the direct speech of a man named Amundh Niclisson:

- (12) huru kan **ek** sitia her f[yre] er **ek** er ein fatik suein [...]
‘how can I live here, for I am a poor boy [...]
(DN I.760, 1437)

The one instance of <jek>, on the other hand, is attributed to a man called Holta Jonson:

- (13) [...] **jek** vil læna tek han [...]
‘[...] I will rent it [the farm] to you [...]
(DN I.760, 1437)

We do not know very much about Amundh and Holta’s background; none of them have been given titles in this charter. However, from the context, it seems clear that Holta has a higher social rank. Amundh is quoted to refer to himself as poor, while Holta is in a position to rent Amundh a farm where Amundh and his wife can live for the rest of their lives. The broken form <jek> thus seems to be associated with a higher social class, consistently with the findings presented in section 2.

3.1.3 Point of view

In some charters the use of broken vs. unbroken pronoun forms correlates with changes in point of view; in these cases not only the form of the 1st sg. nom. pronoun changes, but also its deictic reference. This variation bears some resemblance to direct/reported speech, discussed above in section 3.1.2, but is independent of quotation. A plausible underlying reason for variation correlating with changes in point of view is dictation (see Mørck 1999:271, citing Hægstad 1902:4, Brøndum-Nielsen 1934, Seip 1954a:216 and Seip 1954b:219), in combination with the fact that written charters (especially early documents) *recorded* and only gradually replaced oral legal transactions (Larsson 2001:190). Although we are generally working on the assumption that the first signatory is the best candidate for the person to be considered the ‘speaker’ of a charter (see footnote 6), we can easily imagine scenarios in which not only this signatory, but also other parties exert influence on written forms via their oral statements. We might particularly expect this type of

influence to affect lexical choices (Mørck 1999:271), and the use of broken forms can reasonably be labelled a lexical choice, given our finding that broken forms of the 1st sg. nom. pronoun were introduced via Swedish, and not through a regular phonological process.

3.1.2

A charter in which point of view seems to be relevant is DN XI.216 (1464, Berg [Vestfold]), which has two signatories, Ewind Sigurdzsson and Roar Niclessen. The document has 3 instances of the 1st sg. nom. pronoun, once in the unbroken form <ek> and twice in the broken form <jek>. The form <ek> is used when the point of view is with Roar, as illustrated in Extract 14.

- (14) Sporde **ek** tha Roar lensman, ader^{do} Erlande aat, om han hadhe tha siit granna skæll [...] ‘Then I, sheriff Roar, asked the aforementioned Erland if his neighbours would testify [...]’ (DN XI.216, 1464)

Shortly after this, the point of view changes from Roar to Ewind, and the 1st person sg. nom. pronoun is now spelt <jek>, as shown in Extract 15a–b:

- (15) a. Spurdæ **jek** tha Ewind Sigwrdzsson proffaster j Tonsberge [...] oc sagde swa, Dande men oc dande quinnor som hær ero nw kompna j dag, hwat rykthæ eller hwat granna skæll wilia j bera æder witna j dag [...] ‘Then I, Ewind Sigurdzsson, provost in Tønsberg [...], asked this: ‘good men and good women who have come here today, what truth or what neighbour's testimony do you want to give or bear witness to today [...]’
- b. Sagde **jek** tha Ewinder Sigurdzsson proffaster før^{der}, til men oc quinnor som thær nerwarande, waro, haller oc rækker vp yder hender. alle, som med honum witna wilia j dag [...] ‘Then I, the aforementioned provost Ewind Sigurdzsson, said to the men and women who were present there: ‘raise your hands, everyone who will testify for him today [...]’ (DN XI.216, 1464)

The signatories Roar and Ewind have different titles; Roar is a *lénsmaðr*, while Ewind is a chorister and a provost. In terms of the classification presented in section 2.3.2, this makes Roar a lower official and Ewind a member of the lower clergy. Our quantitative investigation showed that around the time when the charter was issued (1464), the proportion of broken forms was rapidly increasing in documents whose first signatory was a lower official; it was actually somewhat higher than in charters whose first signatory was lower clergy (cf. figure 9). This gives us some reason to expect <ek> from the clergyman and <jek> from the lower official, i.e. the opposite pattern of what we find. However, there is still a non-negligible amount of variation between broken and unbroken forms in both social groups at the relevant time, and the micro-level variation associated with the two individuals Ewind Sigurdzsson and Roar Niclessen is thus not incompatible with the macro-scale patterns of change that were described in section 2. We can only speculate about the reasons for the distribution of the pronoun forms in this specific case; perhaps it reflects a difference in age between Roar and Ewind. At any rate, the pattern of variation in DN XI.216 illustrates how the qualitative, micro-level approach can complement the macro-level approach and provide a more nuanced picture of the contemporary linguistic variation in the medieval communities.

Another charter in which point of view seems to influence the realisation of the 1st sg. nom.

pronoun is DN VI.618 (1493), which records a deal between Laurens Torstensson and Villiam Olaffsson concerning shares in a farm named Tvilde [Hordaland]. The 1st sg. nom. pronoun occurs 4 times, twice in the broken form and twice in the unbroken form.

The first part of the charter reports the sale of the shares. The 1st sg. nom. pronoun only occurs in contexts of direct or reported speech, but, importantly, the quotations are given from the point of view of the signatories (Isaac Anfynsson and three other magistrates). The pronoun forms are broken (<ieg>, <jeg>); cf. Extract 16.

- (16) a. tha sporda for^{de} Laurens Torstensson Villiam Olaffsson vm han var man til ath selliæ thet iordæ gotz man blyffuer **ieg** til ath kaupæ thet
‘Then the aforementioned Laurens Torstensson asked Villiam Olaffsson if he would be able to sell that property, “then I will be able to buy it”’
b. tha svarede Villiam Laurens igen **jeg** vil loffue tek thet [...]
‘then Villiam answered Laurens: ‘I will promise you that [...]’
(DN VI.618, 1493)

Towards the end of the charter, however, the point of view is shifted to Villiam, who states that he has received the payment from Laurens. The 1st sg. nom. pronoun now appears in the unbroken form <egh>, as shown in Extract 17.

- (17) kennis oc **egh** tit nempdher Villiam Olaffson ath **egh** haffuer vp borith aff titnempdher Laurens Torstensson myndræ penningh oc meyre oc aller ther i mellem som i kaup vare kom
‘I, the oft mentioned Villiam Olaffson, also acknowledge that I have received from Laurens Torstensson the full payment as agreed.’
(DN VI.618, 1493)

We propose that the unbroken forms in this passage reflect the language of Villiam Olaffsson.

The variation in DN VI.618 provides an illustration of the difficulties that can be involved in teasing apart factors leading to intra-textual variation. An alternative interpretation could be that instead of point of view, the crucial factor is direct or reported speech. However, the passage attributed to Villiam Olaffson is distinguished by another linguistic feature in addition to the unbroken pronoun form: the word form <meyre> ‘more’ has retained a diphthong which has been monophthongised when the point of view is with the signatories. We see this in the form <hetther> ‘is called’ in the body text preceding Villiam Olaffson’s statement, and also in the form <ytermere> ‘further’, which is a part of the eschatocol that more or less directly follows it.

3.1.4 Co-authorship or time gap?

Charters are commonly assumed to be produced by a single scribe (e.g. Wetås 2008:89). However, in one of the charters in our corpus, variation between broken and unbroken pronoun forms correlates with other characteristics in a way that could possibly be compatible with co-authorship. The charter in question is DN III.739 (1437, Oslo); cf. illustration 1 for a photo of the document.

Illustration 1: DN III.739 (1437, Oslo). Photo: Ann-Sofi Pettersson, The National Archives of Sweden.

In DN III.739, two different parts can be distinguished. In the first part, which includes the first 10 lines and the beginning of line 11, the 1st sg. nom. pronoun is consistently spelt <jak>, and the ink is rather pale and blurry. When unabbreviated, the conjunction *ok* is spelt <oc>, and long vowels are not marked by gemination, except in the word *svá*, spelt <saa>. In the second part, which starts with <Ok> on line 11 and includes the rest of the document, the 1st sg. nom. pronoun is consistently spelt <ek>. The ink is darker and clearer, and the letters generally look slightly bigger than in part 1. When unabbreviated, the conjunction *ok* is spelt <ok>, and long vowels are to some extent marked by gemination. Vowel gemination affects certain words whose vowel length is *not* marked in part 1 of the document; compare <knwt> on line 3 to <knuut> on line 11/12, and <sina> on line 1 to <siin> on line 13.

Although there are many interesting palaeographic differences between part 1 and 2, there is no clear difference in the hand. A possible explanation for this could be that the scribes were from the same scriptorium. However, it is also conceivable that there was in fact only one scribe involved, who was interrupted and finished the charter at a later point, using different forms.

3.2 Intra-individual variation

In section 3.1, we focussed on variation in pronoun forms within individual charters. However, our

annotated corpus also facilitates investigations of variation associated with individual signatories, across charters. In this section we will present two case studies, Hacon Lafranson and Aruidher Ingeldzson.

3.2.1 Case study 1: Hacon Lafranson

Hacon Lafranson was a lawyer in Jämtland and Herjedalen, and is the sole signatory of three different charters dated between 1426 and 1434. In the first document, the 1st sg. nom. pronoun is mostly spelt <ek> (3 occurrences), except in the eschatocol, where an additional witness is called to seal the charter, and the pronoun form is <iak>. The variation is illustrated in Extract 18.

- (18) a. [...] kennist **ek** Hacon Lafranson logman j Iæmptalande ok Hæriadal ath **ek** war aa almenneliko þinge j Suegh sokn [...]
'[...] I, Hacon Lafranson, lawyer in Jämtland and Herjedalen, acknowledge that I was at the general assembly in Sveg parish [...]'
- b. ok þil sannenda bidher **iak** ærlikan man ok wællbornan herra Jon kirkio herra j Suegh at han sit insigle henge met mino insigle [...]
'and to testify this, I ask the honest and well born man Jon, priest in Sveg, to put his seal on this letter alongside mine [...]'
(DN XVI.86, 1426)

In the second charter, DN IX.247 (Forberg, 1428), there is no variation, but the 1st sg. nom. pronoun form is not <ek>, as one might expect given the distribution in DN XVI.86; instead, <iak> is used consistently throughout the document (4 times). The third charter, however, resembles the first charter in that <ek> is most common (though it only occurs twice), except in the eschatocol, where <iak> is used. The charter is DN VI.450, which was also discussed in section 3.1.1 on formulaic contexts; the variation is illustrated in Extract 19.

- (19) a. [...] at **ek** var a Aspasom a logleggha stemno [...]
'that I was at the legal meeting at Aspås [...]'
- b. [...] ok þil sannenda hengir **iak** mit insigle fore þetta bref [...]
'and to testify this, I put my seal on this letter [...]'
(DN VI.450 1434)

To sum up, the charters signed by Hacon Lafranson show both intra-textual variation and variation over time. It is not clear how the consistent use of <iak> in DN XVI.86 should be interpreted, but it is interesting to note that two documents by the same signatory exhibit a similar type of variation, i.e. <iak> in the eschatocol, but <ek> elsewhere. As mentioned in section 3.1.1, this could be due to the use of exemplars or conventions specific to certain formulae.

3.2.2 Case study 2: Aruidher Ingeldzson

Aruidher Ingeldzson is the sole signatory of three charters dating from the period between 1430 and 1434. Interestingly, variation between broken and unbroken pronoun forms is found in all of these documents.

The first charter, DN V.597 (Nidaros [Trondheim], 1430), is Aruidher's will. The document is rather long, and the 1st sg. nom. pronoun predominantly has the form <ek> (25 times); however, <iak> occurs once, in the eschatocol, where two witnesses are asked to seal the document. The variation is illustrated in example 20.

- (20) a. Oc hafuir **ek** gifuit þer till domkirkiune jordh mina Myklaboollstadhe [...] 'And for that I have given my land Myklebostad to the Cathedral [...]'
- b. [...] þa beiddis **ia**k beskeidhna manna Odd Jonssons oc Jwta Knwtsons radmanna i Nidharos incigel medh mino firi þetta testamenta bref [...] ' [...] I ask for the seals of the good men Odd Jonsson and Jwta Knwtson, councillors in Nidaros, along with mine on this will [...]' (DN V.597, 1430)

In the second charter, DN V.609 (1432), in which two councillors from Trondheim act as witnesses, Aruidher gives goods to the choristers in Nidaros. In this document there is more variation than in DN V.597: the unbroken form <ek> is used 5 times, whereas broken forms occur 8 times (spelt <iak> twice and <jak> six times). Some examples are given in Extract 21.

- (21) a. jtem antwardhadhe **ek** oc ther medh optnemdom kanunkom jordhena Petuikena [...] 'Likewise, I gave the previously mentioned choristers the farm Petvik [...]'
- b. Haffwer **ja**k theim vnt oc frelsleka vp latit her fore allan min gardh j Berwen [...] 'For this, I have granted them and freely given up all of my farm in Bergen [...]'
- c. jtem giffwer **ia**k oc fornempdom koorsbrødrum j Nidroos [...] thua lytene j halffwe badhstuffuone som mek til hørðhe vppa berghene i Bergwine [...] 'Likewise, I give the aforementioned choristers in Nidaros [...] my shares in the sauna up on the hills in Bergen [...]' (DN V.609, 1432)

In the third charter, DN V.640, which is another will, <ek> is again the predominantly used form, occurring 14 times. <iak> occurs twice, however, and the distribution differs from that in the first will (DN V.597) in that <iak> is found in the body text, not in the eschatocol. The variation is illustrated in Extract 22.

- (22) a. **E**k Aruidher Jngeldz son heilbrigdho aat minne samuitzsko. en tho at **ek** kranker se till likam mini gerir nw thetta mith testamentum gudhi till lofs oc heidher. 'I, Aruidh Jngeldz son, mentally sound, though I am physically ill, now make this will to honour God.'
- b. [...] oc ther mædh alt annat ther **ia**k aghir som ekki er skift oc bort gifuit. ' [...] and moreover, everything else that I own that has not been distributed and given away.' (DN V.640, 1434)

The use of exemplars probably explains the variation in the first will (DN V.597); we might alternatively suggest that <iak> here reflects the influence of the witnesses. However, we need to

consider additional factors to be able to account for the variation in the other charters. A possibility that presents itself when we examine the background of Aruidher Jngeldzson more closely is to relate the variation to dialect mixture.

In addition to being the first (and sole) signatory of the charters discussed above, Aruidher Jngeldzson is mentioned in various capacities in 13 other charters dated between 1402 and 1439.¹⁵ These charters do not provide any immediately relevant instances of *ek* or *jak*, in some cases because Aruidher is not the first signatory, other times because the 1st sg. nom. pronoun does not occur. However, the documents seem to indicate that Aruidher has moved, a fact which may be reflected in his language use.

As we have seen, Aruidher is the first signatory of multiple charters connected to Trondheim. However, most of the other charters in which Aruidher is mentioned are from Bergen, where he is said to have been a councillor (*ráðsmaðr*). The Bergen charters are older than the Trondheim charters; they are dated between 1402 and 1426. Thus, it looks as if Aruidher originally came from the south-western part of Norway, where we can expect unbroken pronoun forms to be the local norm. Later he moved north to the Trondheim area, where broken forms were introduced at an earlier stage than in the South-West (cf. figure 10). The broken forms in Aruidher's documents from Trondheim may be a result of influence from the linguistic environment there.¹⁶

3.3 Summary

While our quantitative investigation yielded a very clear picture of how the introduction of broken forms proceeded on a macro level, the qualitative approach revealed patterns of micro-level variation that are much more difficult to grasp. We have discussed both intra-textual and intra-individual variation and argued that a number of factors are at work, some of which are more easily discernible than others. First, we made the case that intra-textual variation may arise in contexts of formulaic language, quotations or changes in point of view; we also found one charter in which variation between broken and unbroken forms may be due to co-authorship or a time gap in the production of the document. We then presented two case studies of intra-individual variation, using our annotated corpus to trace how broken vs. unbroken forms are used by individual signatories in different charters over time.

As before, we have arrived at findings that could not have been identified with any other method. Drafting practices, formulae and co-authorship all contribute to variation in a chaotic way which could not be recovered by examining broad statistical trends. Individual variation, both intra- and inter-textual, can speak to individual life histories in a way which population statistics can only hint at. In short, a qualitative approach to these data was required to uncover the full richness of linguistic variation.

¹⁵ The relevant charters are DN XII.136 (1402), DN I.630 (1411), DN II.620 (1412), DN XIV.33 (1417), DN II.641 (1417), DN I.670 (1420), DN I.683 (1422), DN XII.177 (1424), DN II.684 (1425), DN XII.179 (1425), DN XII.184 (1426), DN V.656 (1436) and DN V.686 (1439).

¹⁶ It is worth mentioning that Aruidher also seems to have ties to the areas north of Trondheim: in a charter from 1417 (DN II.641), Aruidher buys a farm in the parish of Hol in Buksnes [Nordland]. He is also one of several witnesses in a charter that was written on behalf of the populace of Hålogaland and Finnmark, although it was published in Bergen (DN I.670, 1420). The parts of Nordland that are included in our maps seem to generally pattern like Trondheim wrt. breaking, but there is too little data north of Trondheim to be confident of the linguistic patterns in these regions; any statements about the dialect of Buksnes in this period would be little more than guesswork.

4 CONCLUSION

In this paper, by studying the same dataset with contrasting methods, we have made the case for a unified quantitative-qualitative approach in historical sociolinguistics. Each view of the dataset has uncovered different findings; each finding enriches and contextualises the others, so that we arrive at a whole that would have been inaccessible by any methodologically simpler study.

There are many stories about breaking in the 1st sg. nom. pronoun to be told. Some social and cultural generalisations emerge at the levels of countries and populations. Our quantitative results led us to conclude that broken forms entered written Norwegian due to Swedish influence, first in the east and in Trøndelag; in eastern Norway, this development was led by female writers and by those from the higher social classes; writers of the middle classes (lawyers, lower officials) preserved local traditions of writing better and so maintained unbroken forms for longer; in western Norway, where the change was purely a written phenomenon never reflected in speech, this social differentiation did not take place.

However, as our qualitative investigation showed, these broad generalisations are not the whole story: we have argued that especially in the central period of high variation, use of broken or unbroken form was a socially embedded choice that served different communicative functions in different moments and places. The incoming broken pronoun could form part of an urban, East Nordic model and, from a practical standpoint, be a sign of using an East Nordic exemplar or East Nordic drafting practices. Conversely, the unbroken form could indicate a closer connection with traditional drafting practices and local norms. As the change progressed and broken forms became more universal, unbroken forms were increasingly relegated to slips of the pen: the local dialect of the scribe or signatory showing through when their grasp of the written standard faltered.

The variation in writing sometimes reflected parallel variation in speech, thus bringing to bear a similar complex of social meaning. The broken forms may have been particularly associated with the speech of the wealthy and powerful, unbroken forms with poorer speakers. Equally, once the broken forms had become established in the dialects of eastern Norway, the use of one form or another might have served as a signal for the dialect of the speaker being quoted or the writer themselves. Alternating between forms could serve to indicate the boundaries of quoted speech or switching perspective and authorship within a text.

Unsurprisingly, this variation existed not only between the speech and writing of different people but also within the production of single individuals over time and in different documents. Such intra-individual variation may have served communicative purposes (selecting one variant or another according to formality, localness, etc.) but could also simply reflect the changing linguistic circumstances of an individual's life (living in communities in different times and places where the change had progressed differently).

What is most striking about the trajectory of the change and data from the period of high variation is the complexity. Many factors interacted, at times with contradictory results. Such mixed phenomena are typical of historical language data and call for mixed and adaptable methods: a purely quantitative, macro-scale investigation would fail to pick up the many fascinating local and

ephemeral patterns in the data, yet conversely a purely qualitative approach would get lost in the detail and miss the big picture. Only by combining these approaches, as we have done here, is it possible to approach a fuller understanding of the whole, messy narrative; we believe that this methodology could be productive in future studies on a range of other changes that have taken place in the history of Norwegian, such as the loss of subject-verb agreement and the loss of case, and changes in other languages.

5 REFERENCES

- Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Gabor Grothendieck & Peter Green, 2016. *Package “lme4.”* <https://cran.r-project.org/package=lme4>. Accessed 2016.
- Berg, Ivar, 2013. Eit seinmellomalderleg skrivemiljø: Nidaros erkesete 1458–1537 [A late medieval scribal community: Nidaros archdiocese 1458–1537]. Trondheim: Norwegian University of Science and Technology Ph.D.
- Berge, Alvdis, 1974. Genitiv av substantiv i mellomnorske diplom frå 1425-1426: Ei jamføring med norrønt [Nouns in the genitive in Middle Norwegian charters from 1425-1426: A comparison with Old Norse]. Universitetet i Bergen MA.
- Blaxter, Tam T. 2017. Speech in space and time: Contact, change and diffusion in medieval Norway. Cambridge: University of Cambridge Ph.D. <https://doi.org/10.17863/CAM.15576>.
- Brøndum-Nielsen, Johs., 1934. Om middelalderlig nedskrivning efter diktat [On writing from dictation in the Middle Ages]. In Andersen, Poul, Aage Hansen and Peter Skautrup (eds.) *Studier tilegnede Verner Dahlerup paa femoghalvfjerdsdaarsdagen den 31. oktober 1934* [Studies dedicated to Verner Dahlerup on his seventy-fifth birthday, 31 October 1934], 63–66. København: Universitetsforlaget i Aarhus.
- Culpeper, Jonathan. 2009. Historical sociopragmatics: An introduction. *Journal of Historical Pragmatics* 10(2). 179–186.
- Evans, Mel. 2013. *The language of Queen Elizabeth I: a sociolinguistic perspective on royal style and identity*. (Publications of the Philological Society 46). Malden, MA: Wiley/Blackwell.
- Farstad, Stein E, 1991. Sætrekk ved vokalismen i brev fra Hedmark 1356-1420 [Characteristics of the vowel system in charters from Hedmark 1356-1420]. Oslo: Universitet i Oslo MA.
- Hansen, Maj-Britt Mosegaard. 2009. The grammaticalization of negation markers in Old and Middle French: A discourse-functional approach. In Maj-Britt Mosegaard Hansen & Jacqueline Visconti (eds.), *Current trends in diachronic pragmatics and semantics*, 227–251. Bingley, UK: Emerald Group Publishing Ltd.
- Hødnebo, Finn, 1971. Om diplomer som kilde for norsk språkhistorie [On charters as a source for the history of the Norwegian language]. In Magerøy, Hallvard and Venås, Kjell., editors, *Mål og namn. Studiar i nordisk mål- og namnegranskning*. Heidersskrift til Olav T. Beito [Language and names. Studies in Scandinavian linguistics and name research. Festschrift to Olav T. Beito], 145–153. Oslo: Universitetsforlaget
- Hamre, Lars, 1972. Innføring i diplomatikk. Førellesningar [Introduction to the study of charters. Lectures]. Oslo: Universitetsforlaget.
- Indrebø, Gustav, 2001. Norsk målsoga [History of the Norwegian language]. Bergen: Norsk bokreidingslag. 2nd edition.
- Johannessen, Janne Bondi, Joel Priestley, Kristin Hagen, Tor Anders Åfarli, & Øystein Alexander Vangsnes. 2009. The Nordic Dialect Corpus – an advanced research tool. In Jokinen, Kristiina and Eckhard Bick (eds.): *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009*, 73–80. Odense: Northern European Association for Language Technology (NEALT)
- Knudsen, Trygve, 1927. Betrakninger over to grupper Vestfold-brev fra 1390-årene [Reflections on two groups of charters from Vestfold from the 1390s]. In *Festschrift til Hjalmar Falk 30. desember 1927: fra elever, venner og kolleger* [Festschrift to Hjalmar Falk 30 December 1927: from students, friends and colleagues], 129–141. Oslo: Aschehoug.
- Hægstad, Marius, 1902. Maalet i dei gamle norske kongebrev [The language of the old Norwegian royal charters]. Kristiania: Dybwad.
- Larsson, Inger, 2001. Svenska medeltidsbrev. Om framväxten av ett offentligt skriftbruk inom

administration, förvaltning och rättsutövning [Swedish medieval charters. On the emergence of public writing conventions in the administration, civil service and judicial practice]. Stockholm: Sällskapet Runica et Mediævalia.

- Mazzon, Gabriella. 2003. Pronouns and Nominal Address in Shakespearean English. In Irma Taavitsainen & Andreas H. Jucker (eds.), *Diachronic perspectives on address term systems*, 223–250. (Pragmatics & beyond N.S., 107). Amsterdam: Benjamins.
- Mørck, Endre, 1980. Passiv i mellomnorske diplom [The passive in Middle Norwegian charters]. Oslo: Universitet i Oslo MA.
- Mørck, Endre, 1999. Sociolinguistic studies on the basis of medieval Norwegian charters. In Ernst Håkon Jahr (ed.), *Language change: advances in historical sociolinguistics*, 263–290. Berlin: Mouton de Gruyter.
- Nakaya, Tomoki, Martin Charlton, Jing Yao, Stewart Fotheringham, Paul Lewis & Chris Brunsdon, 2014. GWR4. https://geodacenter.asu.edu/gwr_software. Last accessed 02/05/2017
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics: Language change in Tudor and Stuart England*. Harlow: Pearson Education.
- Noreen, Adolf. 1904. *Altschwedische Grammatik, mit Einschluss des Altnorwegischen*. Halle, M. Niemeyer.
- Noreen, Adolf, 1970, *Altnordische Grammatik. 1: Altisländische und altnorwegische Grammatik (Laut- und Flexionslehre)*. Unter Berücksichtigung des Urnordischen. M. Niemeyer. 5th edition.
- Pettersen, Egil, 1975. *Språkbrytning i Vest-Norge 1450-1550 [Language change in western Norway 1450–1550]*. Bergen: Universitetsforlaget
- Raumolin-Brunberg, Helena. 2009. Lifespan changes in the language of three early modern gentlemen. In Minna Nevala, Arja Nurmi & Minna Palander-Collin (eds.), *The language of daily life in England (1450–1800)*. Amsterdam: John Benjamins.
- Rindal, Magnus, 1989. Nordnorsk skriftspråk i diplom før 1370 [Written Northern Norwegian in charters from before 1370]. *Maal og Minne*. 129–139.
- Rumpf, Jonas, Simon Pickl, Stephan Elspass & Werner König, 2009. Structural analysis of dialect maps using methods from spatial statistics. *Zeitschrift für Dialektologie und Linguistik* 76(3). 280–308.
- Sandøy, Helge, 1985. *Norsk dialektkunnskap [Norwegian dialectology]*. Oslo: Novus forlag.
- Seip, Didrik Arup, 1954a. Et høyere talemål i middelalderen [A higher form of speech in the Middle Ages]. In Didrik Arup Seip, *Nye studier i norsk språkhistorie [New studies in the history of the Norwegian language]*, 192–217. Oslo: Aschehoug.
- Seip, Didrik Arup, 1954b. Skriftnorm og talemål i Norge i det 13. hundreår [Written norm and speech in the 13th century]. In Didrik Arup Seip, *Nye studier i norsk språkhistorie [New studies in the history of the Norwegian language]*, 218–266. Oslo: Aschehoug.
- Tønnessen, Ragnhild, 1995. Verbalbøying i eit utval Oslo-diplom 1450-1499 [Verbal inflection in a selection of charters from Oslo 1450-1499]. Oslo: Universitetet i Oslo MA.
- Vågslid, Eivind, 1969. Vestfoldmål 1366–99 i 30 brev av bisp Øystein Aslakson i Oslo [Vestfold

language 1366–99 in 30 charters by bishop Øystein Aslakson in Oslo]. Oslo: Universitetsforlaget.

Wetås, Å., 2008, Kasusbortallet i mellomnorsk. Ein komparativ studie av proprialt og appellativisk materiale [The loss of case in Middle Norwegian. A comparative study of proper and common noun data]. Oslo: University of Oslo Ph.D.

Willis, David, 2017. Investigating geospatial models of the diffusion of morphosyntactic innovations: The Welsh strong second-person singular pronoun *chdi*. *Journal of Linguistic Geography* 5(1), 41–66.