Convergence in generic pronouns: Language contact and Faroese mann

Remco Knooihuizen*

Centre for Language and Cognition Groningen, Rijksuniversiteit Groningen

Despite state-driven language policy against Danish linguistic influence, the Faroese language has borrowed the Danish generic pronoun mann ‘one’. As in Danish, this pronoun varies with generically used tú ‘you’. An analysis of the variation in Faroese shows that Faroese tú is used more often than its Danish equivalent du (26% vs. 16.5%) and that although there is extensive inter-individual variation, different linguistic factors (inclusion of the addressee in the referent of the pronoun, use in a conditional construction and verb tense) and social factors (age, gender and peripherality of location; possibly also speech style) constrain this variation in Faroese and in Danish. This suggests that extensive bilingualism has not led to sociolinguistic convergence between Faroese and Danish. Inclusion of sociolinguistic analysis in language contact research can help further our understanding of contact processes and inform language policy.

Keywords: convergence; Danish; Faroese; generic pronouns; language contact

1. Introduction

With an estimated 50,000 to 70,000 speakers (Petersen 2010a, 31), Faroese is the smallest of the North Germanic languages. It is the dominant language in the Faroe Islands, although all speakers of schooling age and over also speak Danish, the other official language in the islands and the language of wider communication within Denmark, to which the Faroe Islands belong administratively. Much of the language-sociological discourse about Faroese, then, is concerned with the societal roles and relative prestige of the two official languages, and linguistic research on Faroese often focuses on lexical and grammatical borrowings from Danish. State language policy, too, has – at least until relatively recently – focused on providing Faroese alternatives for Danisms, as evidenced by usage advice books (e.g. Tausen 1996) and inclusion policies for dictionaries.

In the face of these prohibitive language policies, however, there has been considerable transfer of linguistic material between Faroese and Danish, leading some scholars to talk about convergence between the two languages (Petersen...
This concerns both lexical material as well as morphological items and (morpho)syntactic patterns (cf. Sections 2.2 and 2.3 below). Interestingly, some of the items that are transferred from Danish into Faroese are subject to sociolinguistic variation in Danish, which adds another dimension to the question of linguistic transfer: Is transfer restricted to purely linguistic matter, or may sociolinguistic patterns be transferred along with it? And if, as previous research suggests, the latter is the case, what may be the constraints on transferring a faithful copy of this variation from one variety to another?

This paper aims to contribute to this ongoing debate by analysing variation in the use of the Faroese generic pronouns mann and tú, and comparing the results to the equivalent variation in Danish, which has recently been thoroughly investigated (Jensen 2009a and others). The paper is structured as follows: Section 2 introduces the linguistic situation in the Faroe Islands and gives a brief discussion of linguistic transfer between Faroese and Danish. Sections 3 and 4 review the literature on (socio)linguistic transfer and generic pronouns, respectively. The results of a corpus study into the use of mann and tú in Faroese are presented and discussed in Sections 5 and 6. The conclusion in Section 7 offers comments on the possibility of sociolinguistic transfer, and an outlook for further research.

2. Faroese and Danish in the Faroe Islands

2.1. Social setting

The Faroe Islands, settled from Norway c. AD 800, came under the Danish crown with the Union of Kalmar in 1380. In the following centuries, Danish rule was established in the islands, with Danish appearing as the administrative language of the islands around 1500 (Rasmussen 1987, 55). This was the start of a gradual process culminating in today’s situation of early sequential Faroese-Danish bilingualism (Petersen 2010a, 35).

Historical post-reformation sources – although their reliability can sometimes be questioned, and the precise interpretation of terms used may be unclear – suggest that Danish rapidly became widespread in the Faroe Islands (Petersen 2010a, 37). A great push for Danish came with the introduction of Danish schooling in the Faroe Islands in the mid-nineteenth century (Rasmussen 1987, 48), after which the Faroe Islanders’ Danish is said to have improved and Danish influence on Faroese to have increased (see below). Danish received a further boost from the 1930s onwards, as many elements of modern society came from Denmark, and in Danish.

In the Faroe Islands today, Faroese is clearly the dominant language in most aspects of life, but Danish has a significant presence especially in education and as a reading language. Faroese children acquire (some) Danish before formal acquisition starts in schools at age 10–11 (Petersen 2010a, 40). While education is clearly a factor in maintaining Danish influence, Danish also has a significant presence in the everyday Faroese linguistic landscape (television,
books and magazines; Kühl 2013, 17). Danish proficiency in the Faroe Islands today is good, with students achieving exam results comparable to those of Danish speakers in Denmark (Petersen 2010a, 41).

2.2. Danish influence on Faroese – and back

Centuries of increasing bilingualism with Danish have left their mark on Faroese, resulting not only in many loan words and loan translations, but also in many (possibly) contact-induced changes in Faroese. Lexical loans from Danish are possibly the most widespread, and definitely the most visible in Faroese. Petersen (2010a, 91) shows that 30% of the nouns, verbs and adjectives in his corpus were Danish; unfortunately, he does not discuss his classification criteria. The loans have led to a significant purist backlash (Petersen 2010a, 44–48), but not all of the proposed alternatives are equally well accepted (Clausén 1978, 131–134).\(^1\) Attitudes to loan words and purist alternatives differ, often depending on social factors, but estimates of the proportion of the population in the purist and pragmatic camps vary significantly (Jacobsen 2008, 298). Lexical loans are generally adapted to Faroese phonology and morphology (Clausén 1978, 60–61).

In addition to lexical loans, there is also morphological and syntactic influence from Danish on Faroese (Heycock, Sorace, and Hansen 2010; Petersen 2010a, 99, 2010b), including the borrowing of the generic pronoun mann – the focus of this article. Quite a few of the morphological and syntactic loans involve variation, and although Petersen often gives patterns of variation in Faroese, no comparison is made to patterns of variation in Danish, even if some of the features are variable also there.

Faroese-Danish bilingualism has of course also led to Faroese influence on the Danish that is spoken in the Faroe Islands. The Faroe-Danish\(^2\) of the younger generations today shows limited Faroese interference in phonology, due to increased exposure to models from Denmark (Kühl 2013, 19), but other types of interference do to some extent remain. Elements of Faroese interference on Faroe-Danish may be style dependent: Kühl’s (2011) analysis of high-school essays shows that stylistic constraints on written Faroese are transferred to written Faroe-Danish, resulting in ungrammatical language from a Mainland Danish point of view. In all cases she analysed, the grammatically correct Danish corresponds to the colloquial variant in Faroese, while the formal Faroese variant does not have a Danish counterpart. The bidirectional influence of Faroese and Danish shows how ingrained bilingualism is.

---

\(^1\)My impression of the words used in Clausén’s study is that many of the less accepted words have made significant inroads since the late 1970s, and that they are now frequently used both in speech and writing.

\(^2\)The term gotudanskt is often used for Faroese forms of Danish, but this term is probably best used only for those forms that show considerable Faroese interference in phonology (spelling pronunciation) as well as lexicon, morphology and syntax (Poulsen 1993; Petersen 2010a, 50).
3. Language contact and variation

3.1. Language contact

A contact-induced change, writes Thomason (2003, 688), is any change ‘that would have been unlikely, or at least less likely, to occur outside a specific contact situation’. In many cases, such change is ‘interference’, making the changing language structurally more similar to the language it is in contact with.

Two elements of the Faroese-Danish language contact situation are of interest here. Firstly, widespread bilingualism in the source and receiving languages (terminology from van Coetsem 1988) raises the chances of a contact-induced innovation getting beyond actuation and being picked up by the community, as listeners are more likely to recognise and understand the innovative form and may use it in their own speech; see e.g. Matras (2009, 76) on the social facilitation of collective inter-language features. Secondly, close similarities between the two languages in contact is also thought to be conducive to contact-induced change: innovations are more easily adopted and less disruptive to the receiving language (Thomason 2003, 687). If we assume a cline of similarity between the varieties in contact, as in Siegel (2001, 191), contact-induced changes should be easier between more closely related – or more accurately, more similar – languages. Bowern (2013), however, offers convincing evidence that the degree of contact-induced change attested in a language is related not to the similarity between the languages in contact, but to the type and intensity of speaker contact. As also Siegel (2001, 193) alludes to, contact-induced change is first and foremost a sociolinguistic issue.

3.2. Variation and contact-induced change

Sociolinguistics, of course, is closely linked to the study of language change, and the foundations for modern-day variationist sociolinguistics were laid in a paper explicitly about language change (Weinreich, Labov, and Herzog 1968). But beyond the charting of the propagation of a change and analysing the social meanings of the resulting variation, variationist sociolinguistics is also a useful diagnostic tool for contact-induced change: by comparing constraints on variation in a particular linguistic variable, we may hypothesise about its origin. If the variation is constrained in a (near) identical manner in multiple varieties, one plausible explanation is that it shares the same heritage for all varieties – in other words, it is the same variable that has either spread from one variety to others, or that has arrived in all varieties through regular transmission from an earlier stage of the language. This idea is the basis for comparative sociolinguistics (Tagliamonte 2002); like Labov’s (1972, 120) definition of the speech community, it highlights underlying constraints on variation over actual frequencies of variants.

Variationist sociolinguistics uses multiple logistic regression to analyse linguistic data, but has developed slightly idiosyncratic terminology (Johnson
An independent variable, e.g., gender, is referred to as a ‘factor group’ (statistical terminology: ‘factors’); male and female are ‘factors’ within that factor group (statistical terminology: ‘levels’). Effects are reported as ‘factor weights’ between 0 and 1. A factor weight over 0.5 favours the use of the variant in question; a factor weight below 0.5 disfavours it. The greater the deviation from 0.5, the bigger the effect size of the factor is. Factor weights over and below 0.5 can be converted into positive and negative log odds, respectively.

The variationist method has previously been used in cross-dialectal and cross-linguistic studies of (contact-induced) language change. Among these are studies comparing the constraints of variation on varieties of the same language across geographical space (Buchstaller 2008; Buchstaller and D’Arcy 2009), across time (Poplack and Malvar 2006), or both (Naro and Scherre 2000); and by comparing speakers of the same variety with different degrees of exposure to different varieties, it has been used in studies of dialect maintenance and change (Meyerhoff and Walker 2007, 2013). As a tool for researching language contact-induced change, there have been comparisons of constraints on variation in the two languages of bilingual speakers (Auger and Villeneuve 2008; Blondeau and Nagy 2008), and in two languages that stand in a superstrate/substrate relation to each other (Meyerhoff 2009). It is this last approach that is most relevant to the present study, where the L1 of a bilingual group (Faroese) is compared to a monolingual L2 group (Danish).

Meyerhoff (2009, 303) and Buchstaller and D’Arcy (2009, 316) propose that the type, or strength, of transfer is determined by the faithfulness with which constraints on variation have transferred along with the linguistic form; a schematic overview can be found in Table 1. The weakest form of transfer in this classification is ‘surface transfer’, when only a linguistic form is transferred from the source to the receiving language, but none of the associated constraints on variation have. In other words, there is superficial similarity between the corresponding elements in the two languages, but they work in different ways. Meyerhoff does not include this level of transfer in her overview, however, as linguistic form need not be relevant in language contact: in cases of pattern replication (Matras 2009, 234–274), there may be transfer of function without transfer of form.

Table 1. Criteria from comparative variationist sociolinguistics for different transfer types in language and dialect contact, from Buchstaller and D’Arcy (2009, 316).

<table>
<thead>
<tr>
<th>Criteria</th>
<th>None</th>
<th>Surface</th>
<th>Weak</th>
<th>Strong</th>
<th>Calque</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Significance of constraints</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ranking of constraints</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Hierarchies of constraints</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
</tbody>
</table>
Three further levels of transfer are defined by the extent to which constraints on variation have been copied. In ‘weak transfer’, the same factor groups are significant in constraining the variation, but the relative importance of the factor groups or the way in which they influence the variation may differ. ‘Strong transfer’ means that not only are the same factor groups significantly constraining the variation in both languages, but also that the relative order of the factor groups is identical. ‘Calquing’ builds on this by also requiring the factors within each of the relevant factor groups to be ranked identically.

In order to classify a linguistic innovation as contact induced, however, a socio-historically plausible contact situation and similar constraints on the variation in the putative source and receiving languages are not enough, especially when dealing with pattern replication where the linguistic form has not been transferred. Similarities between constraints may also arise because of factors other than contact. In Meyerhoff’s (2009, 311–312) study of null arguments in Tamambo and Bislama (Vanuatu), one of the shared constraints was the presence or absence of the argument in the previous clause. This shared factor can easily be explained by discourse salience; such quasi-universal factors, that can be expected to operate independently of the individual language, do not offer good enough evidence of contact-induced change.

4. Generic pronouns
4.1. Generically on generic pronouns
The variable investigated in this study is the use of the generic pronouns tú and mann in Faroese. A generic pronoun – also ‘impersonal’ or ‘indefinite’ pronoun – is a pronoun that refers to no single particular person. The referent of a generic pronoun is obligatorily human; variation exists between languages in whether the referent of a generic pronoun may or may not (have to) include the speaker and/or addressee of the utterance (Hoekstra 2010, 43).

Frequent grammaticalisation sources of generic pronouns are words for ‘man’ (German, Danish man; Dutch, Frisian men; multi-ethnic London English man, Cheshire 2013; Icelandic maður, Heine and Kuteva 2002, 208; cf. also Brazilian Portuguese a gente ‘people’, Zilles 2005) or ‘one’ (German ein, Danish en, English one; Heine and Kuteva 2002, 221).³ In addition to these forms, the second-person singular pronoun can also be used with generic reference. An early Scandinavian example is given in (1), from the fourteenth-century Gotlandslagen (Haberland 1986, 82):

³Icelandic maður is good evidence in support of Faroese mann being a Danish loan: had this been independent grammaticalisation, the nominative maður would have been a stronger candidate for a subject pronoun than the apparent accusative mann. While tokens of maður as a generic pronoun can be found sporadically in the 1890s periodical Føringatiðindi, perhaps modelled on Icelandic usage, the form is not used in present-day Faroese.
Variation in the choice of generic pronouns has been particularly well described for Danish, which the Faroese data in this article will be compared to and which will be reviewed separately in Section 4.3, and for French (Laberge and Sankoff 1980; Ashby 1992; Coveney 2003; Blondeau 2008). The various studies discuss a range of factors constraining the variation between generic *on* ‘one’ and *tu* ‘you’ in different varieties of French, with a small number of syntactic and pragmatic factors regularly appearing as significant. *On* rather than *tu* is favoured after so-called ‘presentative heads’ such as *il me semble que* ‘it seems to me that’ and *je trouve que* ‘I think that’ (Ashby 1992, 145). Both appear in conditional constructions, both the antecedent and the consequent clause, but conditionals with *si* ‘if’ favour *on* whereas conditionals with *quand* ‘when’ favour *tu* (Ashby 1992, 145). Finally, when the statement offers moral evaluation, *on* is favoured, while *tu* is favoured in situational insertions (Laberge and Sankoff 1980, 280–285; Ashby 1992, 148; Blondeau 2008, 263).

### 4.2. Generic pronouns in Faroese

The three Faroese generic pronouns included in this study are *mann*, *tú* and *ein*, corresponding to the Danish pronouns *man*, *du* and *en*. This section discusses and exemplifies the various uses of these pronouns in Faroese; the Danish pronouns are discussed in the following section.

The first of the Faroese pronouns is *mann*, often – wrongly, according to *Føroysk orðabók* – spelled *man*. *Mann* is a Danish loan in Faroese and is generally stigmatised in formal and written discourse. *Føroysk orðabók* tags the word as ‘spoken language’, and it is absent from the 1961 Faroese-Danish dictionary, which takes a relatively purist stance on the Faroese lexicon. Similarly, *mann* is completely absent from grammars (Andreasen and Dahl 1997; Thráinsson et al. 2004) and language courses (Lockwood 1977; Henriksen 1983; Adams and Petersen 2009; Petersen and Adams 2009). Tausen (1996, 64–66), in a published collection of style advice for broadcast media, writes that ‘This word *man* has held little regard in writing, and many try to avoid saying *man*, because we have

\[\text{(1)} \text{drepr } \mathbf{\text{bu}} \text{ mann ... } \mathbf{\text{pa}} \text{ byt } \mathbf{\text{briar marcr; slar } \mathbf{\text{bu}} \text{ mih stangu epa} \text{ kill you man ... then pay three marks, hit you with stick or} \]
\[\text{yxar hambri, byt siex oyra, sargar } \mathbf{\text{bu}} \text{ mann, byt tolf oyra axe.gen hammer, pay six øre, wound you man, pay twelve øre} \text{ ‘If you kill a man ... , then pay three marks; if you hit with the handle or} \]
\[\text{the head of an axe, pay six øre; if you wound a man, pay twelve øre’}.^4
\]

^4Unless otherwise stated, translations of non-English material in this article are my own.
so much other good language to use instead’. What follows is an extensive list of alternatives, including the pronouns ein, tú and teir/tey ‘they’; the nouns fólk ‘people’, maður ‘man’ and menn ‘men’; a range of impersonal constructions in the passive or middle voice; or descriptive strategies.

The main use of mann is as a generic pronoun, as in (2). Alternatively, it may have clear first-person singular reference, as in (3). Sporadic cases are found where mann has first-person plural reference. Unlike the other personal pronouns, mann does not have any non-nominative forms.

(2) viss mann skal hava parkeringspláss so skal mann vera øgiliga heldigur
‘If one must have a parking space, one has to be very lucky’.

(3) sjálvandi svimjihøllin bleiv bygd eisini tá ið mann var smádrongur
‘Of course the swimming pool was also built when one was a little boy or something, right’.

The primary referent of tú, the second-person singular personal pronoun, is of course the addressee of the utterance, as in (4). As in many other languages, it may also have generic reference, as in (5). In addition to these two pronoun uses, tú may also be used as a discourse marker, initiating a topic (6), or as part of the discourse marker tú veit ‘you know’ (7).

(4) so eg havi sæð teg uttanfyri har eg tonkti tú sá mær so kend út
‘So I have seen you outside there. I thought you looked so familiar to me’.

(5) ja ja men tú fær ikki hondverkara-útbúgving viss tú gongur eitt
‘Yeah, but you don’t get a vocational diploma if you only go one year, but if you go those two years, then you do’.

5All Faroese examples are taken from the Nordic Dialect Corpus, which is described in 5.1.

6The discourse marker tú veit ‘you know’ may also appear as tú veistst or, with inversion, veit(st) tú. Faroese has dialectal variation in the use of the second-person verb ending -st, but uninflected tú veit predominates regardless of geographical area (Knooihuizen 2014, 97–98).
The final generic pronoun in this study is *ein ‘one’, the first alternative to *mann* given in Tausen (1996, 64), from which (8) is taken. Note that in this conditional construction, the uptake of *ein* is * hann ‘he’, in contrast to (5), where both forms are *tú*.

4.3. Generic pronouns in Danish

The Danish pronouns *man*, *du* and *en* have very similar functions to their Faroese counterparts discussed above (cf. Jensen 2009a). They will therefore not be exemplified in detail here. In Danish as in Faroese, *man* only occurs as a subject pronoun; although the distribution of pronouns shows a more complicated pattern, *en* is, in a sense, used as the object form of *man*. The discourse-particle function of *du*, exemplified in (9), is slightly different from that of Faroese *tú* in (6): it has an intensifying function rather than an initiating one. Finally, although also Danish has *du* ved ‘you know’, Jensen (2009a) does not discuss this use of *du*.

(9) *ha der kan jeg huske da fik jeg altså en bagi *du*
ha there can I remember then got I really one in.behind you
*så jeg kunne mærke det ikke*
so I could feel it not
‘At that time, I remember, I really had my bottom smacked, you bet, so I could feel it, right’ (Jensen 2009a, 88)
et al. 2013). These studies form an excellent comparison for the current study of Faroese. As the details of the Danish variation are best discussed in parallel with the Faroese data, the discussion here will focus only on the larger patterns in the linguistic and social constraints.

The strongest linguistic constraint on the use of man is that it is exclusively used as in subject position. If we were to take en as the object form of man, however, a subject constraint continues to operate: the proportion of man in subject position is much higher than that of en in object position, where du is much more frequent (Jensen 2009a, 94). Du is used more frequently in cases when the hearer is included among the referents of the pronoun (cf. below) and in conditional constructions (Jensen 2009a, 107, 109). Traditional social factors receive relatively little attention in the Danish studies, but it was found that working class speakers used more du than middle class speakers, and that men used more du than women (Jensen 2009a, 104).

In addition to such underlying constraints on man and du, the choice of generic pronoun is also determined by usage-based, interactional constraints. There is a tendency for du to appear in talk about the present, and in the present tense, while man tends to appear when the past is discussed, often with the past tense (Nielsen, Fosgerau, and Jensen 2009, 123). Du further tends to be used to illustrate or exemplify a claim, to ‘show’, while man is used to ‘tell’. Using du, then, signals a higher degree of involvement from the speaker or invites the interlocutor to involve (Nielsen, Fosgerau, and Jensen 2009, 126, 129, 132).

A remarkable aspect of the man ~ du variation in Danish is its geographical patterning, and the shift in these patterns over time. Generic du in written sources is at least attested from the eighteenth century (Jensen 2009a, 90). In spoken language, Maegaard et al. (2013, 15–19) observed a sudden rise and subsequent decline in usage rates in the period from 1978 to 2010. Du spread from Copenhagen to the rest of Denmark, but appears to have peaked in the capital already in the 1980s, before other areas caught up with the rise. Rates of du use started to fall in Copenhagen while they were still on the rise in the other locations in the study. After a later peak, these places now appear to be catching up with the general decline of du (Jensen 2009a, 98; Maegaard et al. 2013, 18–19). The exception to this pattern appears to be Vinderup in Jutland, the most peripheral location in the study seen from Copenhagen, where du usage is particularly high. Part of the explanation may lie in the delay of the rise-fall pattern, but it is also noteworthy that the alternative, man, is not used in the traditional dialect in Vinderup, so that du in a sense faces less strong opposition (Jensen 2009b, 167).

4.4. Classification

In order to analyse the effects of the referent of the generic pronoun on the choice of a specific pronoun man or du, Jensen (2009a, 105) classified the pronouns into
four groups. This applies only to those instances of man and du that have not previously been classified as having specific (first-, second- or third-person) reference or as discourse particles. Jensen’s four groups are as follows:

- everybody or a group of people not further defined but including both speaker and addressee
- a group of people including the speaker, but excluding the addressee
- a group of people including the addressee, but excluding the speaker
- a group of people excluding both the speaker and the addressee.

‘In the overwhelming majority of cases’, writes Jensen (2009a, 106), ‘it was fairly straightforward to categorise the pronouns into one of [these] four types’, although sometimes the classification relied on taking into account a large section of discourse context. In very few cases, where no decision could be made, the token was excluded from further analysis. While classifying the Faroese data according to this system, however, I found a considerable amount of tokens where the classification was not clear cut. Consider, for example, (10), in which it is unclear from the discourse context whether the proposition also applies to the addressee (we do not know if he uses hay to feed his sheep); or (11), where mann may be a non-generic pronoun with first-person reference (i.e. the interviewer asking the question) or a generic pronoun with a referent including the speaker and perhaps the addressee too (the more general interpretation given in the translation).

(10) hevur mann ikki hoyrt um nógyvan seyð í Hvalba?
    has one not heard about much sheep in Hvalba?
    ‘Isn’t Hvalba known for having lots of sheep?’

(11) so slapst tú frá hasum rukkulinum við hoyggi
    then escaped you from that hard.knock.life with hay
    ‘… then you wouldn’t have to faff about with hay anymore’

In order to facilitate coding, I restructured Jensen’s classification into a feature-based system with four binary dimensions. The dimensions used are also at the basis of Jensen’s classification; the system can account not only for Jensen’s four types of generic pronouns, but also for the other pronominal uses of man/mann and du/tú.

- Does the pronoun have specific or generic reference?
- Does the referent of the pronoun include the speaker of the utterance?
- Does the referent of the pronoun include the addressee of the utterance?
- Is the referent of the pronoun a (larger) group of people?
5. Corpus study

5.1. The corpus

The investigation of current use of generic pronouns in spoken Faroese is based on the Faroese material in the Nordic Dialect Corpus (Johannessen 2009; Johannessen et al. 2009). This part of the Nordic Dialect Corpus contains just over 6 hours of speech (approximately 45,000 words) by 20 speakers from 5 locations in the Faroe Islands (Figure 1). Speech is equally divided over two speech styles: conversations (one-on-one dyads between speakers matched for location, age group and gender who often knew each other in advance) and interviews (one-on-one dyads between a speaker and one of two female interviewers from Tórshavn). The corpus was collected in 2008 as part of a larger data collection project on Faroese and transcribed by a native speaker. Although the fieldwork aimed for an even spread of gender and age groups across locations, there is a slight imbalance in the corpus (Table 2).

5.2. Data and methods

All occurrences of mann, tú and ein were extracted from the corpus and coded for whether they were used as a generic pronoun. Given that mann can only be used as a subject, non-nominative forms of tú and ein were not included in

Figure 1. Map of the Faroe Islands with the five locations surveyed in the Nordic Dialect Corpus.
the analysis as they lie outside the envelope of variation. The analysis here builds on only those tokens for which generic reference was clear; ambiguous tokens were rejected. This left 413 tokens of generic mann, 143 tokens of generic tú and no tokens of generic ein. We are left with a binary choice of generic pronouns: mann or tú.

The data were then coded for a number of factors concerning the referent (the inclusion of speaker, addressee and/or others; see above), social factors (speaker, speaker age, speaker gender, location and speech style), and linguistic factors (verb tense and the use in a conditional construction).

5.3. Historical baseline

In order to contextualise the picture of language change in Faroese beyond the apparent-time results of the corpus study, a small analysis was done of the use of the pronoun mann in the early Faroese periodical Føringatíðindi (1890–1901, 1906). This is one of the oldest available text corpora of any reasonable size, consisting of material from different authors, and although this is written language which, in addition, may not be entirely representative of late nineteenth-century Faroese due to the publication playing an important role in linguistic nation-building at the time, the analysis may in principle give a baseline against which to compare the results from the modern-day spoken study.

A total of 119 uses of generic mann and tú were found in all issues of Føringatíðindi together, divided over the years as in Table 3. (The publication frequency differs between years, and the number of issues for each year is given in the table. Each four-page issue consists of 2250–2500 words of prose text. Poetry, recurring advertisements and incidental material in Danish are excluded.) The overall proportion of mann among generic pronouns is 43%,

\begin{table}[h]
\centering
\begin{tabular}{|l|cc|cc|}
\hline
Location & Younger speakers & Older speakers \\
& Male & Female & Male & Female \\
\hline
Fuglafjørður & 0 & 2 & 2 & 0 \\
Klaksvík & 0 & 2 & 2 & 0 \\
Sandur & 2 & 1 & 0 & 1 \\
Tórshavn & 2 & 0 & 0 & 2 \\
Tvøroyri & 0 & 0 & 2 & 2 \\
\hline
\end{tabular}
\caption{Number of speakers, by age group and gender, for each of the five locations surveyed in the Faroese part of the Nordic Dialect Corpus.}
\end{table}

\footnote{The alternatives listed in Section 4.2 (generic tey, fólk and maður) did not occur in the corpus.}

\footnote{All issues of Føringatíðindi are available online at \url{http://www.timarit.is} in facsimile and in plain text with (uncorrected) optical character recognition.}
but there are significant differences between years, and what is immediately
striking is the spike in occurrences of generic mann, both in relative and
absolute terms, in 1896. A possible explanation may lie in the fact that the
publication has had different editors, who must have followed a different pol-
icy with regard to this feature. Although there are many more tokens in 1896,
the uses of generic mann in that year do not appear to differ qualitatively from
the other years; all are similar to examples (2) and (3) above.

All in all, what the Føringatíðindi data shows is that generic mann is not a new
development in Faroese. Its use, however, appears to have been avoided by editors
for purist purposes, which leaves us with an unreliable baseline for comparison.

5.4. Results

Of the total of 556 generic pronouns in the data, 26% were tú and 74% were
mann. This mirrors the Danish situation, which shows 16.5% du and 83.5%
man, in addition to a negligible number of tokens of en in subject position
(Jensen 2009a, 94).

The overall picture, however, hides wide-ranging intra-individual variation
(Figure 2), with two speakers – at either end of the age range – making exclu-
sive use of mann, and four speakers using 50% or more tú. The big intra-indi-
vidual differences may be reminiscent of personal-pattern variation, where
pronoun choice would not be constrained by social or linguistic factors, but
purely by the individual speaker (Dorian 1994, 2010), but upon closer inspec-
tion, the variation does appear to show such constraints.
A multivariate analysis in Rbrul (Johnson 2009), including speaker as a random factor, selected two factor groups as significant: speaker gender, and an interaction between speech style and inclusion of the hearer in the referent of the pronoun (Table 4 and discussion below). Other factors were not significant. In the following, I compare these results to the constraints on the man ~ du variation in Danish. Note that the Danish material (Jensen 2009a, 2009b; Maegaard et al. 2013) focuses on du rather than man – an inverse focus compared to the main interest in this article – and that the Danish multivariate analysis (Maegaard et al. 2013, 33–34) uses different methods; in particular, their analysis uses treatment contrasts as opposed to Rbrul’s sum contrasts (see Johnson 2009, 361). The methods are not directly mutually convertible; Table 5 gives an approximate summary of the Danish results in a format compatible with Rbrul output. To facilitate comparison with the Faroese study, the order of factors within factor groups has been reversed to show constraints on the use of man rather than du. Because of Maegaard et al.’s use of treatment contrasts, factor order only applies when all other factor groups are kept at their

Speaker is included in the analysis as a random factor to account for individual preferences of speakers that may distort estimates of the influence of speaker characteristics such as age and gender (Johnson 2009, 363–365).
the baseline levels, which are underlined in the table. (On different methods for sociolinguistic analysis, see Tagliamonte 2012, 138–157).

5.5. Social factors

The Danish data show evidence of language change in apparent time, with the proportion of du decreasing from 22% in the oldest age group to 8% in the youngest age group and the proportion of man increasing accordingly. When

Table 4. Results of the variable rule analysis for mann (vs. tú) in Faroese. Total N = 556; input probability = 0.791; deviance = 473.263.

<table>
<thead>
<tr>
<th>Factor</th>
<th>n</th>
<th>% mann</th>
<th>Factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (random factor)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Style × Referent includes hearer (p = 2.64 × 10^-8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview × hearer excluded</td>
<td>180</td>
<td>91</td>
<td>0.781</td>
</tr>
<tr>
<td>Conversation × hearer included</td>
<td>160</td>
<td>66</td>
<td>0.439</td>
</tr>
<tr>
<td>Conversation × hearer excluded</td>
<td>105</td>
<td>66</td>
<td>0.402</td>
</tr>
<tr>
<td>Interview × hearer included</td>
<td>111</td>
<td>67</td>
<td>0.348</td>
</tr>
<tr>
<td>Gender (p = 0.00176)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>308</td>
<td>87</td>
<td>0.698</td>
</tr>
<tr>
<td>Male</td>
<td>248</td>
<td>59</td>
<td>0.302</td>
</tr>
<tr>
<td>Town</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tórshavn</td>
<td>107</td>
<td>77</td>
<td>[0.684]</td>
</tr>
<tr>
<td>Fuglafjørður</td>
<td>100</td>
<td>78</td>
<td>[0.564]</td>
</tr>
<tr>
<td>Klaksvík</td>
<td>144</td>
<td>74</td>
<td>[0.467]</td>
</tr>
<tr>
<td>Tvøroyri</td>
<td>107</td>
<td>70</td>
<td>[0.406]</td>
</tr>
<tr>
<td>Sandur</td>
<td>98</td>
<td>72</td>
<td>[0.374]</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pensioner (b. 1934–1942)</td>
<td>38</td>
<td>87</td>
<td>[0.632]</td>
</tr>
<tr>
<td>Older (b. 1953–1966)</td>
<td>251</td>
<td>73</td>
<td>[0.476]</td>
</tr>
<tr>
<td>Younger (b. 1979–1992)</td>
<td>267</td>
<td>73</td>
<td>[0.391]</td>
</tr>
<tr>
<td>Verb tense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past tense</td>
<td>107</td>
<td>90</td>
<td>[0.589]</td>
</tr>
<tr>
<td>No verb</td>
<td>18</td>
<td>83</td>
<td>[0.498]</td>
</tr>
<tr>
<td>Present</td>
<td>431</td>
<td>70</td>
<td>[0.413]</td>
</tr>
<tr>
<td>Conditional construction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>376</td>
<td>77</td>
<td>[0.561]</td>
</tr>
<tr>
<td>Consequent</td>
<td>58</td>
<td>71</td>
<td>[0.473]</td>
</tr>
<tr>
<td>Antecedent</td>
<td>122</td>
<td>67</td>
<td>[0.465]</td>
</tr>
<tr>
<td>Referent includes speaker</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker included</td>
<td>474</td>
<td>73</td>
<td>[0.507]</td>
</tr>
<tr>
<td>Speaker excluded</td>
<td>82</td>
<td>81</td>
<td>[0.493]</td>
</tr>
</tbody>
</table>
the Faroese data is divided into three age groups separated by natural large gaps between the informants, the opposite picture emerges: the oldest age group uses considerably more *mann* than the middle and younger age groups. Note, however, that the middle and younger age groups perform exactly like the data overall, and that it is the older age group – two individuals with a total of only 38 tokens – that deviates. In contrast to the Danish data, speaker age is not a significant predictor for the use of *mann* in Faroese. In other words, the apparent-time picture is unlikely to be meaningful.

A much clearer differentiation is found when the informants are separated according to speaker gender (or, more accurately, speaker sex). As can also be glanced from Figure 2, the proportion of *mann* use is much higher among female speakers than among male speakers; in fact, speaker gender is one of the two significant predictors of *mann* use in Faroese. The gender pattern is similar to that in Danish, albeit much more extreme: for Danish *du*, usage rates were found of 14% for female speakers and 18% for male speakers, and speaker gender was found to be a significant predictor only in the more recent data (Maegaard et al. 2013, 16, 33).

There are small geographical differences in the data: the preference for *mann* over *tú* appears strongest in the more centrally located places, while the use of *tú* is higher in more peripheral locations. This factor is not significant.

Table 5. Predictors of the use of generic *man* (vs. *du*) in Danish, adapted from Maegaard et al. (2013, 34). Factor groups are given in order of significance; factors within factor groups are given in order of decreasing preference for *man*. Non-significant factor groups are included in square brackets. Baseline levels in the original model are underlined.

<table>
<thead>
<tr>
<th>Factor group</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker (random factor)</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Yes &gt; No (continuous; younger speakers use more <em>man</em>)</td>
</tr>
<tr>
<td>Year of Birth</td>
<td>No &gt; Yes</td>
</tr>
<tr>
<td>Conditional</td>
<td></td>
</tr>
<tr>
<td>Locality</td>
<td>Vinderup &gt; Odder &gt; Næstved &gt; Copenhagen</td>
</tr>
<tr>
<td>Time of recording</td>
<td>2000s &gt; 1980s</td>
</tr>
<tr>
<td>Referent includes hearer</td>
<td></td>
</tr>
<tr>
<td>[Social class]</td>
<td>No &gt; Yes</td>
</tr>
<tr>
<td>[Gender]</td>
<td>MC &gt; WC</td>
</tr>
<tr>
<td></td>
<td>Female &gt; Male</td>
</tr>
</tbody>
</table>

Significant interactions

<table>
<thead>
<tr>
<th>Location × Time of Recording</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of Recording × Social Class</td>
</tr>
<tr>
<td>Conditional × Time of Recording</td>
</tr>
<tr>
<td>Location × Referent includes hearer</td>
</tr>
<tr>
<td>Subject × Location</td>
</tr>
</tbody>
</table>
however. Geographical patterns of change in Denmark are more complicated: initially, an increase in the use of *du* spread from the undisputed centre Copenhagen to the detriment of other generic pronouns, but as peripheral areas in Denmark are still catching up with that change, a new change decreasing the use of *du* is now spreading from Copenhagen, partly obscuring a centre/periphery pattern in the Danish data (Maegaard et al. 2013, 17 [Figure 4]). Overall, though, locality is a significant predictor in Danish, with central places generally still using more *du* than peripheral locations.

Finally, there are clear stylistic differences in the data, with *mann* being used much more frequently in interviews (82%) than in conversations (66%). Although the Danish studies do not report on style effects, this factor is still interesting for an investigation of sociolinguistic transfer as the findings run counter to the generalisation that Danisms (in this case, *mann*) tend to be avoided in high-register speech. However, I argue that the style difference may in fact not be genuine; see the discussion for more details.

### 5.6. Linguistic factors

Three language-internal constraints were found to influence the choice of *man* or *du* in the Danish data; these were also investigated for Faroese. Firstly, like Danish *du*, *tú* is found relatively more often in conditional constructions. This factor is among the most significant predictors in Danish, but is not significant in Faroese. Then, the Danish studies found that *du* is used to talk about states or events in the present using present tense while *man* is used to talk about the past in the past tense (not included in the multivariate analysis); the Faroese data shows a similar pattern in that *tú* is used more frequently if the verb is in the present tense (including present perfect) than if the verb is in the past tense. Again, this factor is not a significant predictor in the Faroese data.

The third language-internal constraint is that of the inclusion of the hearer in the referent of the pronoun. This is a significant predictor in Danish: if the referent of the pronoun includes the hearer, there is a slight preference for *du*, if it does not, there is a slight preference for *man*. In Faroese, this constraint interacts with style to become the most significant predictor of the choice between *mann* and *tú*. The distribution of generic pronouns that do or do not include the addressee in the referent is very different in conversations compared to interviews (Table 6; chi-square test, $p < 0.001$). A possible explanation for this is

<table>
<thead>
<tr>
<th>Referent includes hearer</th>
<th>$n$</th>
<th>Yes (%)</th>
<th>No (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversation</td>
<td>265</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>Interview</td>
<td>291</td>
<td>38</td>
<td>62</td>
</tr>
</tbody>
</table>

Table 6. Use of generic pronouns that include the addressee in the referent, by speech style.
the result of pragmatics and the dyadic nature of the conversations, rather than any genuine stylistic variation. Whereas in interviews, participants are explicitly requested to talk about themselves – and note that another function of *mann* is as a first-person pronoun – participants in a conversation are much more inclined to engage their interlocutor, for example by including them explicitly in the referent of generic pronouns by using *tú*. In order to still be able to investigate the influence of speech style and the inclusion of the hearer in the referent, these two factors were replaced by an ‘interaction group’ in the analysis: all four possible combinations of style and hearer inclusion were analysed as separate factors. It is this interaction group that is the most significant constraint on the variation in the model. The results show that hearer inclusion in the referent is not a factor of importance in conversations (the relevant factor weights are very close together), but that it does matter in interviews (where the factor weights are very far apart). In interviews, the inclusion of the hearer in the referent works as in Danish: if the hearer is included, there is a preference for *tú*.

6. Discussion

The presentation of the results in the previous section already showed many similarities and differences between the Faroese data and the findings on the Danish *man* ~ *du* variation. In this section, I first discuss some of the notable results before moving on to a discussion of the role of Faroese-Danish language contact in the constraints of this variable.

A first striking finding is that not a single token of generic *ein* was used in the corpus, even though *ein* is a preferred variant in language advice. (Of course, the advice applies especially to written, not spoken, registers). That the advice is otherwise taken heed of is apparent from the fact that in the WiFDF corpus of written texts in Faroe-Danish, there is in fact a marked overuse of Danish *en* where mainland Standard Danish would prefer *man*, precisely because written Faroese requires *ein* (personal communication, Karoline Kühl). This development may be linked to the evolving Faroese-internal diglossia alluded to earlier, as the notion that apparent Danisms are to be avoided in formal styles is carried over to speakers’ second language, in this case Danish.

Evidence of whether the *mann* ~ *tú* variable in Faroese is currently undergoing change is difficult to interpret. The apparent-time method suggests no change in progress: although grouping the data into three age groups appears to show a preference for *mann* among the oldest speakers and stable variation (slightly preferring *tú*) in the other groups, the age groups gloss over considerable intra-group variation, and speaker age was not selected as a significant predictor in the multivariate analysis. The earlier (written) data from Føringatíðindi suggest *tú* was the preferred variant in the 1890s; a mismatch with the apparent preference for *mann* among older speakers. The overall real- and apparent-time picture is difficult to match to the Danish data, where after an initial rise before the 1980s, use of *du* has recently started to decrease.
Previous work on the Faroese material in the Nordic Dialect Corpus found no difference in language use between conversations and interviews, a finding that was explained by the fact that the close-knit social networks in the Faroe Islands did not necessitate any kind of social self-positioning of the speaker by means of language (Knooihuizen 2014, 101). It is therefore surprising that the use of generic pronouns does differ considerably between the two speech styles. The mann ~ tú variation again shows the difficult position of prescriptive language advice: in interviews, where speakers are supposedly more aware of their own speech, the social incentive to engage the interlocutor using tú is absent, causing mann to be more frequent. In the less self-conscious conversational speech, this incentive does exist, and paradoxically, the stigmatised variant mann is therefore less frequent in non-policed than in policed speech.

A comparison of the constraints on the mann ~ tú variation in Faroese to those on the man ~ do variation in Danish is not unproblematic as the two corpora differ in structure, and some factors were included in one analysis that could not be included in the other. However, comparing the influence of the constraints that are included in both studies shows some major differences: gender is a significant constraint in Faroese, but not in all of the Danish data, and conversely, speaker age, locality and the use of the pronoun in a conditional construction are significant constraints in Danish, but not in Faroese. Only the inclusion of the hearer in the referent of the pronoun is a significant constraint in both languages – but in Faroese, the constraint only applies to interviews, not conversations.

By comparing the Faroese and Danish data to generic pronoun variation in other languages, we can speculate to what extent this similarity is due to language contact or is affected by more universal patterns of variation. Assuming that the social constraints on this variation are language and situation specific, it is the linguistic factors that may be subject to universals; unfortunately, however, the large amount of research done on French on and tu (Section 4.1) has predominantly looked at other constraining factors. The inclusion of the addressee in the referent and the tense of the verb were not taken into account in the French studies, factors such as presentative heads and discursive effects did not feature in the Danish (and, therefore, the Faroese) studies. The only factor to be investigated in all studies was the occurrence of the generic pronoun in a conditional construction, but here the French data shows a split between the conjunctions si and quand. The influence of conjunction choice was not investigated in Danish, and it would be difficult to find specific correspondences between conjunctions in languages that are not very closely related. As a common universal factor, then, this leaves only the preference for tú/du (and possibly tu) in cases where the referent includes the hearer; given the other use of this form as a second-person pronoun, it simply makes sense.
7. Conclusion

This article has discussed the variation in the use of the Faroese generic pronouns *mann*, *tú* and *ein* in the light of possible sociolinguistic transfer from Danish in spite of language policy and attitudes that favour Faroese alternatives over Danish or Danish-influenced forms. Analysis of a small corpus of spoken Faroese showed that the prescriptively preferred form *ein* was not used at all; and different social and linguistic constraints were relevant to the variation between *mann* and *tú* when compared to Danish *man* and *du*. The only fully shared constraint can be argued to be discourse related and possibly universal. In terms of the levels of (sociolinguistic) transfer set out in Section 3.2, then, the mismatch between constraints on the *mann ~ tú* variation in Danish and Faroese suggests this is a case of surface transfer only. Despite extensive bilingualism in the Faroe Islands, the use of the stigmatised ‘Danish’ generic pronoun *mann* does not point at sociolinguistic convergence of Faroese and Danish.

Constraints on variation in varieties in contact have previously been compared for dialects of the same language, and also for different languages, whether they are closely related (Picard and French, Auger and Villeneuve 2008) or not (French and English, Blondeau 2008). The setting investigated in this article adds a new type to this list: two closely related languages of which one (Danish) is extraterritorial. To what extent extraterritoriality – or rather, degrees and types of exposure to the contact language – is a relevant factor, remains to be seen in future research. A clearer picture may emerge, for example, when the Faroese data is not only compared to Danish data from Denmark, but also to Faroe-Danish data (preferably from the same speakers, although naturalistic Danish language conversation between speakers of L1 Faroese may be difficult to come by for sociolinguistic reasons). The triangulation of results may also give an indication of the precise route of the influence and the role that Faroe-Danish plays in any Danicisation of Faroese. An awareness of this role can inform language policy and engender a more nuanced discussion of language and bilingualism in the Faroe Islands.

Acknowledgements

I thank Janne Bondi Johannessen and the team at Tekstlaboratoriet at the University of Oslo for giving me the chance to work with the Faroese data in the Nordic Dialect Corpus. Members of the LANCHART centre at the University of Copenhagen, in particular Torben Juel Jensen, have given useful comments on earlier incarnations of this work, as have Nanna Haug Hilton and Naomi Nagy. The article has also benefited from critical commentary by the editors and two anonymous reviewers for Acta Linguistica Hafnia- sia. All shortcomings, of course, remain my own.

Disclosure statement

No potential conflict of interest was reported by the author.
Notes on contributor
Remco Knooihuizen is a lecturer in English Language and Linguistics at Rijksuniversiteit Groningen and studies sociolinguistic variation and language contact in English and Scandinavian languages at the Centre for Language and Cognition Groningen.

References


