Variation in Infinitive Markers in American Norwegian

Åshild Søfteland, Michael T. Putnam, and Arnstein Hjelde

1. Introduction

The purpose of this incipient work is to take a closer look at observed variations within infinitive markers in American Norwegian (AmNo). AmNo data from the CANS corpus (Johannessen 2015) reveal that ‘te’ occasionally appears as an optional infinitive marker, whereas ‘å’ continues to be the most common in both AmNo and contemporary Norway Norwegian (NoNo) (exemplified here with written Norwegian ‘Bokmål’ in examples (1) and (2)):

(1)  
vi lære **te** bruk skrivemaskin  
vi lærte **å** bruke skrivemaskin  
we learnt to use writing-machine

‘We learnt how to use a typewriter.’

(2)  
ja dette (.) e **te** gLøme  
ja dette det kommer jeg aldri **til å** glømme  
yes this this come I never to forget

‘Yes, this I will never forget.’

The wordform /te/ (in written Norwegian ‘til’) can also surface as a preposition, cf. English ‘to’.

Example (1) and (2) show two somewhat different variants, in that the uttered ‘te’ would correspond to written ‘å’ (infinitive marker) in (1) and to ‘til å’ (preposition + infinitive marker) in (2). Therefore, in spite of the fact that the majority of examples of infinitive markers in the CANS corpus reflect what is usually found in NoNo, there still exists a substantial number of ‘non-standard’ forms.

There are at least three possible explanations for the use of te as an infinitive marker in AmNo, listed here as Hypotheses 1–3:

**H1:** Te as an AmNo infinitive marker is influenced by the English infinitive marker ‘to’ either phonetically and/or systemic: the infinitive marker ‘to’ in English is typically unstressed and, as such, pronunciations as [tǝ] and [te] would be the expected result of a phonological integration into AmNo (Haugen 1953: 427-428, 434). It is also possible to conjecture that the English pattern, where the infinitive marker and preposition have the same phonological form, has served as a model for restructuring in AmNo. Seen from a diachronic point of view, we can describe this as /å/ ⇒ /te/.

Where we now find te, there used to be å in the early stages of AmNo.

**H2:** Te as an infinitive marker is a heritage dialect feature from areas in Norway where this is or has been a part of the local dialect (cf. Aasen 1848: 135 etc.). Seen from a diachronic point of view, we can describe this as /å/ ⇒ /te/. Where we now find te, there also used to be te in the early stages of AmNo, thus marking a retention and (maybe) a (hyper)extension of a dialectal form.

---

1. Saskatoon-SK-14gk, CoonValley-WI-10gm etc. are the ‘informant codes’ from the corpus, informing about where the recording was made (place and state), age group, and gender (gk = older woman; gm = older man).

**H3**: *Te* as an infinitive marker is the result of the quite frequent general combination of */te å/ ‘til+å’ (preposition + infinitive marker) in Norwegian (cf. Faarlund 2003), within different constructions. Seen from a diachronic point of view, we can describe this as */te å/ ⇒ */te/. What used to be *te å* in the early stages of AmNo, is now only *te*, and this is also spreading to contexts that used to be only *å* by analogy.

Following the widely-held position that contact-induced syntactic change is highly conservative in nature (Matras 2009), we hypothesize that the variation found in AmNo-infinitive markers relies primarily on the recycling of existing properties of AmNo-grammar, thus providing evidence from integration of aspects of bilingual grammars (Aboh 2015; Putnam et al. 2018). An additional potential guiding factor is the theoretical notion of structural salience (see Polinsky 2018), according to which functional elements occupy the left periphery of derivational units known as phases. For the phenomenon discussed here, this would mean that infinitive markers undergo a shift from appearing as heads in lower structural positions (such as v or T) to occupying positions in the CP-layer of clauses. As demonstrated in Section 6, this phenomenon is not unique to AmNo, but is attested in other varieties of heritage/contact German(ic). In this article, we attempt to gain a better understanding of the forces that shape the variation of infinitive markers in AmNo.

The rest of the article is structured into six sections. First we describe the spoken language corpora from which most of the examples were extracted; CANS – Corpus of American Nordic Speech and LIA Norwegian – Corpus of Old Dialect Recordings (Section 2). Then, we go more into detail about the AmNo data (Section 3) and compare it with older NoNo data (Section 4). Following this section, there is a discussion of whether or not *te* as an infinitive marker could be analyzed as a dialect heritage feature for the CANS speakers (Section 5), which also includes data from Haugen (1953). Section 6 consists of comparative and more theoretical analyses of the described data, and Section 7 sums up with a brief conclusion and directions for future work.

### 2. Corpus Data – Background Information

#### 2.1. America Norwegian Data

As previously mentioned, most of the AmNo data that we build our analyses on come from CANS – Corpus of American Nordic Speech (Johannessen 2015). The main data collection for the corpus was done between 2010–2016, but it also includes older recordings. The recorded material is spontaneous speech (conversations and interviews), and is transcribed (phonetically and orthographically), lemmatized, and tagged for morpho-syntactic features (see corpus webpage for details).

The data collection for this study has been done with CANS v.2 (early fall 2019). This version included the speech of 163 speakers with Norwegian heritage, 100 men and 61 women (two not tagged for gender), reaching from 1st to 5th generations of speakers (most being 2nd to 4th generation). The total number of tokens in the Norwegian speaking part of this version of the corpus is 660,707, but there is a lot of variation in the length of each recording and (hence) the number of tokens for each speaker. The second version of CANS (v.2) included five recordings from 1942, five from 1987–1992, and 153 from 2010–2016.

#### 2.2. Norway Norwegian Data

The LIA Norwegian– Corpus of Old Dialect Recordings (‘LIA’) contains data from speakers born as early as the 1860s, many around 1900. The total number of tokens in this corpus is 3,481,547, based on recordings of 1,347 speakers from 222 places all over Norway. Just like CANS, LIA has phonetic and orthographic transcriptions (aligned), and is lemmatized and morpho-syntactically tagged.

When using LIA data in comparison with AmNo data, we only look at recordings from speakers born earlier than 1920 (506 speakers/1,656,448 tokens) (i.e., speakers born during the years of mass emigration from Norway). It is our judgement that these recordings represent a variety of Norwegian dialects which are best in line with how actual emigrants spoke.

For both AmNo and NoNo data, examples are presented here in the phonetic transcription from the corpora; that is, a transcription with a regular Latin alphabet that tries to depict the actual speech as accurately as possible (see corpus websites for detailed transcription guidelines).
3. America Norwegian Data

We conducted simple searches in CANS (v.2) for infinitive markers directly followed by an infinitive. The main results are reported here:

\[ \text{å + INF} \quad \text{ca. 1,500 examples} \\
\text{te + INF} \quad \text{ca. 300 examples} \]

What we categorize as the infinitive marker ‘te’ in this article also includes other variant pronunciations. Actually, we have found examples of almost all the possible vowels; \(/\text{ta\}/ /\text{te\}/ /\text{ti\}/ /\text{to\}/ /\text{tu\}/ /\text{t\æ\}/ /\text{t\ø\}/ /\text{t\å\}/\), in addition to some cases of the non-reduced /til/ and /tel/. One example is transcribed as only ‘t’. Still, /te/ is by far the most frequent.

There are three to four main types of speakers in the corpus:

A. Those who only use \(\text{å}\) as an infinitive marker \(\text{ca. 40\%}\)
B. Those who mostly use \(\text{å}\), but have some examples of \(\text{te}\) \(\text{ca. 40\%}\)
C. Those who have about the same amount of \(\text{å}\) and \(\text{te}\) \(\text{ca. 15\%}\)
D. Those who have mostly \(\text{te}\) or only \(\text{te}\) \(\text{ca. 5\%}\)

Many speakers only produce a few examples of \(\text{te}\), but some use it frequently:

(3) \text{trennt tre fire pRass e prøvde te kämm oppåver her} \quad \text{(CoonValley-WI-17gm)}
omeotre tre fire plasser jeg prøvde å komme oppover her \quad \text{(written Norw. BM)}
about three four places I tried to come upwards here
‘There was about three, four places where I tried to find my way up here.’

(4) \text{men e hadde litt vannsklihet te ferstå dømm} \quad \text{(CoonValley-WI-17gm)}
men jeg hadde litt vanskelighet \text{med å} forstå dem \quad \text{(written Norw. BM)}
but I had some difficulty to understand them
‘But I had some difficulties understanding them.’

(5) \text{hænn i læRRde te snakk engels?} \quad \text{(CoonValley-WI-17gm)}
\text{hvor jeg lærte å snakke engelsk?} \quad \text{(written Norw. BM)}
where I learned to speak English?
‘Where did I learn to speak English?’

Some of the frequent users appear to have grammars that diverges more from baseline norms than average, but this requires further investigation. Examples of both ‘å’ and ‘te’ from the same speaker are listed in (6) and (7):

(6) a. \text{så d eR haRt fåR mei te snakke nåssjk} \quad \text{(Saskatoon-SK-02gm)}
så det er hardt for meg å snakke norsk \quad \text{(written Norw. BM)}
so it is hard for me to speak Norwegian
‘So it’s hard for me to speak Norwegian.’

b. \text{så vi må prøve å snakke tibakers i nåRsk} \quad \text{(Saskatoon-SK-02gm)}
så vi må prøve å snakke tilbake på norsk \quad \text{(written Norw. BM)}
so we must try to speak back in Norwegian
‘So we must try to answer in Norwegian.’
(7) a. vi takka dæmm førre re att remm læRde uss te snakke nåRsk (Sunburg-MN-03gm)
we thanked them for that that they learned us to speak Norw.
‘We thanked them for learning us how to speak Norwegian.’

b. jeg tror det blir mer og mer ‘even’ litt ‘grammar’ som som
I think it becomes more and more even some grammar who who

lika å (.) å snakke nokko engelsk å jø kje re? (Sunburg-MN-03gm)
likes to to speak some English too does it?
‘I think it becomes more and more even a little bit of grammar,
who/that likes to speak some English as well, doesn’t it?’

c. det var så morsomt å snakke med alle demn (Sunburg-MN-03gm2)
It was so funny to speak with all of them
‘It was so much fun to speak with all of them.’

Some speakers also produce the combination /te å/ ‘til å’, when written Norwegian Bokmål would have only å.²

(8) a. jeg måtte lære meg te å taLa enngl’st (CoonValley-WI-31gk)
I had to learn how to speak English.

b. det er ikke beste plassen å reise på kRus (SpringGrove-MN-05gm)
it is not best place to to travel on cruise
‘That is not the best place to go for a cruise.’

c. de snakka inte (.) hadde inngen te å snakke norsk med (SpringGrove-MN-05gm)
they spoke not had no-one to to speak Norw. to
‘They did not speak, (they) had no-one to speak Norwegian to.’

² (7c) shows a methodological problem. The transcription says ‘å’ here, but actually it is impossible to hear. Because of the last word before ending in -t and ‘å’ often being pronounced with more of a schwa-sound, we cannot say whether the -t only belongs to the adjective ‘morsomt’ or to the infinitive marker as well. In other cases like this in the corpus, it seems to be transcribed as ‘å’ most of the time, so one could say that the number of possible ‘te’ infinitive markers is higher than listed.
³ In CANS we also find examples of /te å/ or only /tel/ used instead of other prepositions, often targeted as ‘med å’ or ‘for å’. If we do not specify searches for phonetic ‘å’, we also find examples of ‘å’ = /asså/ ‘og så’.
For the CANS speakers, we have some metadata stating the geographical background of the speakers’ Norwegian ancestors. A majority of the speakers have a mixed (or unclear) background; their Norwegian families often originate from several different parts of Norway. Still, we know two things for certain. First, in the communities where some Norwegian is still spoken, the speakers, in general, have a predominantly Eastern Norwegian dialect background (e.g., Johannessen & Laake 2012). Second, in CANS the infinitive marker *te* is used by speakers with ancestors from all main dialect areas in Norway.

### 4. Norway Norwegian Data

As previously stated, *å* is clearly the dominant infinitive marker in modern NoNo, but *te* is also used in the dialects in some areas (cf. Strand 2015; Huus 2018). Older NoNo data, preferably prior to the time of mass emigration to America, is essential in order to achieve an ecologically valid comparison with observed AmNo forms. *Norsk ordbok* (n.d.) mentions infinitive marker *te* briefly in their article on the preposition *til*. Western Norway is the only geographical area that is mentioned in the article. Nesse (2001) also classifies infinitive marker *te* as a Western Norwegian phenomenon, supposedly having spread from Hansa-Bergen to adjacent areas. Aasen (1848: 135/163), based on national field work, mentions both *å*, *til å* and *te* as infinitive markers in Norwegian dialects. He finds *te* to be so frequent that he considers it for the new written language norm, shown in Aasen (1853). In this book, Aasen gives examples of infinitive marker *te* in texts from all main dialect areas.

(9) and (10) show two examples with *te* from the LIA corpus (Western and Eastern Norwegian, respectively):

(9) å so va de te kämma innat å sekkja seg ved rokken (Aurland)
og så var det å komme innatt og sette seg ved rokken
and so was it to come in-again and sit REFL by the-spinning-wheel

‘And then it was just to come back in and sit down by the spinning wheel.’

(10) å så va re nå så vitt n hadde ti te lesa leksusd’n sine (Ål)
og så var det nå så vidt en hadde tid til å lese leksene sine
and then was it now barely one had time to read homework REFL

‘And then we barely had time to do our homework.’

Table 1 shows the result from a search in LIA for *te* (phonetic transcription) directly followed by a verb tagged as an infinitival form—grouped after counties (abbreviated in parentheses). Incorrectly tagged examples have been excluded. Ambiguous examples are included.

<table>
<thead>
<tr>
<th>Geographic area</th>
<th><strong>te</strong> + INF</th>
<th>Examples per 10,000 tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Norway (Ro, Ho, S-F, M-R)</td>
<td>312</td>
<td>7.94 per 10,000 tokens</td>
</tr>
<tr>
<td>‘Mountain Eastern Norway’ (Te, Bu, Op)</td>
<td>51</td>
<td>4.02 per 10,000 tokens</td>
</tr>
<tr>
<td>‘Central Eastern Norway’ (Øs, Ve, Ak, He)</td>
<td>24</td>
<td>1.46 per 10,000 tokens</td>
</tr>
<tr>
<td>Southern Norway (V-A, A-A)</td>
<td>31</td>
<td>1.41 per 10,000 tokens</td>
</tr>
<tr>
<td>Trøndelag (N-T, S-T)</td>
<td>3</td>
<td>0.22 per 10,000 tokens</td>
</tr>
<tr>
<td>Northern Norway (Fi, Tr, No)</td>
<td>7</td>
<td>0.12 per 10,000 tokens</td>
</tr>
</tbody>
</table>

---

4 This analysis is rejected by Faarlund (2003).
5 Whether or not the combination *[til + å]* could be analyzed as an infinitive marker in modern NoNo remains a question for further investigation, but it seems to be possible for some AmNo examples, like those in (8a–c).
6 The preposition ‘*til*’ is also found as an infinitive marker in Swedish, both historically and in modern dialects (Svenska akademiens ordbok (n.d.); Kalm 2016).
As can be seen from the table, te before infinitives is most frequent in the recordings from Western Norway. It appears to some extent in southern and eastern Norwegian, especially in the border areas towards western Norwegian (‘Mountain Eastern Norway’). However, it is infrequent in recordings from the middle and northern parts of the country (‘Trøndelag’ and ‘Northern Norway’).

Furthermore, a closer look at these data from LIA indicates that [te + INF] where written Norwegian Bokmål has [‘til å’ + INF] (as in examples (10) and (2)) was fairly widespread, whereas [te + INF] where Bokmål has [‘å’ + INF] (as in examples (9) and (1)) to a much higher degree was a predominately western Norwegian feature, even if it was also used in some eastern and southern parts of Norway.7

5. Tracing the Developmental Path of Infinitive Marker te

The examples from the CANS corpus analyzed for this paper are mostly newer data. To trace the feature back to its potential roots, we will also take a look at older AmNo data. Examples from Haugen (1953) confirm variation in infinitive markers at an earlier stage (te in (11a-b), å in (11c)):

Speaker: Woman, 83 years old, Spring Grove (heritage: Voss, Western Norwegian):

(11) a. han va so te saia nabogut med oss
    han var så å si nabogutt med oss
    he was so to say neighbor-boy with us
    ‘He was practically our neighbor.’

    b. so at krytri hadde maira kjens te komma igjøno
    slik at krøtteret hadde mer sjanse til å komme igjennom
    so that the-cattle had more chance to come through
    ‘So that the cattle had a better chance to get through.’

    c. men glømt å teke ned atte alt det tåget
    men glømt å ta ned igjen alt det tåget
    but forgotten to take down again all that twig
    ‘But forgotten to take all the twig down again.’

When looking at all the transcriptions in Haugen (1953), we see that the speakers’ geographical background in Norway is spread throughout Mid-/West-/East-Norway (Map 1), but speakers using te as an infinitive marker all have Western heritage (Map 2):

---

7 Data from Huus (2018), a master’s thesis analyzing infinitive markers in the Norwegian part of The Nordic Dialect Corpus (recordings 2006–2012), show quite similar patterns for modern NoNo; te for ‘il’ is less widespread and mostly Western, but there seems to be a general decline in the frequency of te over time if we compare the data in Huus (2018) with our search in LIA.
Map 1: All speakers (Haugen 1953)  Map 2: Speakers using /te/ (Haugen 1953)

Thus, the infinitive marker *te* may have been primarily a western Norwegian heritage dialect feature at an earlier stage. This geographic distribution can be applied to the investigation of the possible origin of these different infinitive markers in AmNo.

Many CANS speakers most likely grew up in communities in America consisting of dialect speakers from various regions of Norway, resulting in ‘dialect mixing’. In CANS, we also find examples of infinitive marker *te* (for *å*) from speakers with no reported western Norwegian background. Most CANS speakers using the infinitive marker *te* also use *å*. Speakers who frequently use the infinitive marker *te* commonly produce other structures that are distinguishable and divergent from expected baselines. As a result, a preliminary conclusion must be that the appearance of *te* as an infinitive marker can at the very least no longer be considered purely a sign of the retention of a vestigial dialectal feature.

**6. Comparative and Theoretical Analyses**

The structural position of these various infinitive markers may also be the result of an underlying tendency in heritage grammars to place functional morphology at the edge of designated structures (commonly referred to as *phases*) in order to make them more ‘salient’. Polinsky (2018) refers to this phenomenon as *structural salience*. Applying this to infinitive markers in Norwegian, this would mean that both infinitive markers *te* and *å* occupy C rather than T, which is illustrated in (12) below (see also Åfarli & Eide 2003: 165):

\[
\text{(12) } \left[ \text{CP C-} \text{te/å TP } \ldots \right]
\]

This hypothesis is supported by additional evidence found in two other varieties of contact German(ic); namely, Brazilian Pomeranian (13; Postma 2018) and Pennsylvania Dutch (14; Börjars & Burridge 2011; Louden 2019), respectively.

**Brazilian Pomeranian**

\[
\text{(13) } \text{Wij arbeira upm laand } \left[ \text{CP taum da arme luur helpen}. \right. \\
\text{We work on the land } \text{for.to the poor people help.INF}
\]

‘We work on the land to help the poor people.’
Pennsylvania Dutch

(14) Ich bin verschtaunt [cf. fer dich sehne do.]
    I am astonished for you see.INF here
    ‘I am astonished to see you here.’

More detailed research is required to substantiate whether (or to what degree) infinitive markers in AmNo have undergone restructuring to achieve structural saliency (i.e., placing the infinite marker in C).

7. Conclusion and Future Directions

The primary goal of this article was to achieve a more detailed understanding of the origin, distribution, and possible motivations between the competing infinitive markers in AmNo. Revisiting the hypotheses introduced earlier (H1-H3), we contend that a combination of these factors has likely contributed to the current state of affairs observed in the CANS corpus. It is unlikely that the emergence of te-infinitives is solely tied directly to a particular NoNo dialect region (H1) (te has been mostly a western Norwegian feature and AmNo societies have been predominantly eastern Norwegian, cf. §4-5). Nor can we completely rule out English influence (i.e., facilitative transfer; H2). Additionally, we cannot ignore the possibility that the t(e) in /te å/ (‘til å’) constructions could be continually affecting what happens to the infinitive marker itself (cf. Faarlund 2003 for older stages of NoNo dialects) (H3). The infinitive marker å is the most frequent in NoNo today, but te used to be more frequent. An additional factor in establishing a definitive (sole) culprit behind the rise of te infinitive markers in AmNo can be found in the nature of immigration itself, during which much inter- and intra-variation from various NoNo dialects could have reasonably competed with each other. The use of te as an infinitive marker could be both language contact induced and dialect contact induced, comprising what the two influencing varieties have in common: the pronunciation [t- + a stressless vowel] for an entity that functions as both a preposition and an infinitive marker.

At this juncture, this study has exposed a very interesting empirical domain that requires more research guided by theoretical grounding. A more fine-grained theoretical analysis of the underlying structural properties of the infinitive markers te and å in AmNo, in comparison with the contemporary written standards (Faarlund 2015; 2019) and dialectal forms, will help refine the research questions and CANS searches. In relation to the first point, a more detailed analysis of the intra-individual switching between te and å for different verbs (cf. Strand 2015) could provide additional insight. Third, sociolinguistic information, especially with respect to background information on those who most frequently produce te-infinitives, will shed light on other extra-linguistic factors that may contribute to their distribution. In summary, the variation of infinitive markers in AmNo represents an interesting domain for inquiry in the years to come.

References


8 The reviewer asks if we have found examples of loans of (other) English prepositions into AmNo. We have not seen that so far, but there are examples of Norwegian prepositions used according to English patterns. For instance, in Hjelde (1992: 29, our translation): De var ifra nord av Trondheim til sør av Oslo ‘They were from North of Trondheim to South of Oslo’—were AmNo lavl (sounding quite similar to English ‘of’) would have been ‘for’ in written Norwegian Bokmål.