Compositional Definiteness in Heritage Norwegian: Production Studied in a Translation Experiment

Yvonne van Baal

1. Introduction

Norwegian in the US has a long research tradition (see Johannessen & Salmons 2012 for an overview). Since 2010, under the ‘Norwegian in America’ project at the University of Oslo, much data have been collected and made available in the Corpus of American Nordic Speech (CANS, Johannessen (2015a)) to document and study American heritage Norwegian. Most of the data collection has been spontaneous speech, although recently experimental data have been collected as well (e.g. Rødvand 2017).

This paper presents data from such an experimental study. The experiment investigates the performance of heritage speakers of Norwegian in the US on the phenomenon Compositional Definiteness (CD). CD is found in Norwegian NPs, where definiteness is marked with a suffixed article on the noun (1a), but with both a suffix and a prenominal determiner when this noun is modified (1b). The lack of either the suffix or the prenominal determiner is ungrammatical (1c-d).

\[(1) \quad \begin{align*}
\text{a. } & \text{bil-en} \\
& \text{car-M.DEF.SG} \\
& \text{‘the car’}
\end{align*}
\]

\[(1) \quad \begin{align*}
\text{b. } & \text{den} \quad \text{fin-e} \quad \text{bil-en} \\
& \text{DEF.SG} \quad \text{nice-DEF} \quad \text{car-M.DEF.SG} \\
& \text{‘the nice car’}
\end{align*}
\]

\[(1) \quad \begin{align*}
\text{c. } & \ast \quad \text{fin-e} \quad \text{bil-en} \\
& \text{nice-DEF} \quad \text{car-M.DEF.SG}
\end{align*}
\]

\[(1) \quad \begin{align*}
\text{d. } & \ast \text{den} \quad \text{fin-e} \quad \text{bil} \\
& \text{DEF.SG} \quad \text{nice-DEF} \quad \text{car}
\end{align*}
\]

The results of the experiment presented in this paper show that with respect to CD, the production of American Norwegian speakers differs in systematic ways from homeland Norwegian. Although there is much variation among the speakers, several patterns can be found. I will claim that most of the patterns that do not coincide with homeland Norwegian cannot be attributed to transfer from English.

\[\ast \text{ Yvonne van Baal, University of Oslo, y.w.h.v.baal@iln.uio.no. Thanks to the audience at WILA 8, Mike Putnam, an anonymous reviewer, Julian Lysvik and Linn Iren Rødvand for their helpful comments. I also want to thank Janne B. Johannessen, Arnstein Hjelde, Kari Kinn and Alexander K. Lykke for their help and support during the fieldwork, Eirik Tengesdal for his help with data transcription, and Ida Larsson for valuable discussions. Special thanks to the American Norwegian speakers who participated in this study.}
\]

\[\ast \text{ As Putnam, Kupisch & Pascual y Cabo (2018) describe, the label ‘heritage speaker’ covers several groups of speakers. The population at focus in this paper belongs to the group of elderly moribund speakers: individuals who are the final generation of speakers with (high) proficiency in the language. This variety of Norwegian is typically referred to as American Norwegian (AmNo) in the literature, and I will use this label throughout the paper.}
\]

\[\ast \text{ The phenomenon is often called ‘double definiteness’. I will follow Anderssen (2012) in using the term Compositional Definiteness instead.}
\]
The paper is structured as follows. Section 2 provides more background on Norwegian NPs and on American Norwegian. Section 3 describes the methodology used and Section 4 shows the results. Section 5 provides discussion and conclusion.

2. Background

The examples in (1) above illustrate the phenomenon of compositional definiteness (CD). However, the system is a little more complex than shown in these examples. First, both the suffix and the prenominal determiner inflect for gender and number (e.g. masculine singular in (1)). Second, there is large variation amongst Norwegian dialects with respect to the morphological form of the two exponents of definiteness. This dialectal variation has to be taken into account when establishing a point of comparison, especially since the speakers of AmNo did not have input from the standard language that could have led to uniformity (Johannessen & Laake 2015: 320). Despite the variation found in the form of the definiteness morphemes, CD is (to the best of my knowledge) obligatory in all dialects of Norwegian.

The structure of Norwegian NPs, and Scandinavian NPs in general, has been studied extensively to account for the complex NPs found. Since Taraldsen (1990) it has been a more or less uncontroversial assumption that Scandinavian NPs display two determiner positions: one below the adjective, associated with the suffixed article, and one above the adjective, associated with the prenominal determiner. Building on this assumption, Julien (2005) provides an in-depth analysis of the Scandinavian DP and shows how the two determiner projections each make their own contribution to the definite interpretation of the phrase.\footnote{Due to space restrictions, the exact syntactic and semantic analysis of CD cannot be discussed here. I refer to Julien (2005) for the complete in-depth analysis. It has to be noted that not all analyses assume the two determiner projections, see Börjars, Harries & Vincent (2016) for an alternative analysis.}

The monolingual acquisition of CD has been studied by Anderssen (2007), showing that although the suffix is acquired very early by Norwegian children, they need more time to acquire the prenominal determiner (ibid: 264). At age 3;3, when Anderssen’s data collection ends, the prenominal determiner is not yet fully acquired, but the suffix is. To my knowledge, there are no studies where the exact age of acquisition of CD is established.

CD has also been studied in speakers of AmNo. Johannessen (2015b: 66) argues in a case study of one speaker that she shows signs of attrition with respect to CD. Anderssen, Lundquist & Westergaard (2018/fc.) study the data available in CANS and show that there is a correlation between CD and possessive structures: speakers who prefer post-nominal possessives, typically leave out the prenominal determiner in CD constructions, whereas speakers who prefer pre-nominal possessives more often leave out the suffix in CD constructions.

Both Johannessen (2015b) and Anderssen, Lundquist & Westergaard (2018/fc.) investigate CD in spontaneous speech. Unfortunately, modified definite phrases are not very frequent (Dahl 2015: 121) and as a result, the numbers per participant are often low. Therefore, the main goal of this paper is to complement these natural speech data with experimental data, in which higher numbers of modified definite phrases in a more controlled setting are elicited. Since non-target-like utterances are found in the spontaneous speech data, it is predicted that these will also be found in the experimental data.

The second aim of this paper is to investigate which element in CD is most vulnerable in AmNo: the prenominal determiner or the suffixed article. This does not have to be the same for all speakers, as shown by Anderssen, Lundquist & Westergaard (2018/fc.).

3. Methodology

The data for this study were collected during a fieldwork trip to the Midwest of the US in fall 2016. The group of researchers that participated recorded speech of 60 speakers in total, of which 19 were analysed for the present paper.

The characteristics of this group of heritage speakers require specific experimental designs. They are elderly speakers, typically not literate in Norwegian, often insecure about their own abilities to speak Norwegian and in general reluctant towards the idea of being tested. Therefore, experiments should be
designed such that they have the least experimental atmosphere as possible, can be administered orally and have simple and short instructions.

3.1. Translation task

To study the production of modified NPs, two experiments were designed: a translation task and a picture naming task. This paper focuses on the results of the translation task only.

In the experiment, participants were asked to translate sentences from their dominant language (English) to the heritage language (Norwegian). The experiment is a collaboration project of three researchers, all studying different linguistic constructions. This means that the test items for the current study were the fillers for their studies, and vice versa. By combining several studies in one experiment, the number of experiments for participants is reduced, which makes a recording session shorter and less demanding.

In order to make participating more enjoyable, the task was presented as a “story telling game”: the participants were told that the researcher would read a story and the participant would help to tell the story in Norwegian by translating it. The story had a total length of 71 sentences. The task started with a training phase of three relatively easy sentences, in which the participant was made familiar with the procedure. At the end, two sentences were used to finish the story. Responses to the training phase and the final two sentences have not been analysed.

The procedure of the task was as follows. The researcher read the sentences of the story aloud one at the time to the participant, who then responded with a translation of the sentence in Norwegian. The whole task was conducted orally, so both the English prompt and the Norwegian response were spoken only. Since the experiment aims to test linguistic competence rather than memory, participants could hear the English sentence as often as necessary.

The nouns used as test items for the present study had to meet two criteria: they had to fit in the story and had to be known by (most) speakers. The latter was met by selecting lexical items from the CANS, i.e. items produced by speakers from the same population. This implied that they were likely to be known by the speakers. Nouns in all three genders were selected as lexical items, as were plural nouns.

In total, the task contains 14 different nouns used in 64 noun phrases. These 64 test items are divided over four context types: indefinite NPs (n=10), definite NPs (n=35), modified indefinite NPs (n=5) and modified definite NPs (n=14). The latter are contexts where CD is required, the other contexts are control conditions. To illustrate, an example of each context is given below.

(2) a. Simple indefinite (control)
English prompt: “flowers”
Target: blomst-er
flower-INDEF.PL

b. Simple definite (control)
English prompt: “the kitchen”
Target: kjøkken-et
kitchen-N.DEF.SG

c. Modified indefinite (control)
English prompt: “a little girl”
Target: ei lit-a jente
F.INDEF.SG little-F girl

4 The other two researchers were Alexander K. Lykke, studying word order and verbal inflection, and Kari Kinn, studying the use of bare noun phrases and kinship terms. During the fieldwork, all three researchers conducted the experiment with different speakers. The recordings were shared among all researchers.

5 There is some overlap between the speakers in CANS and the participants of the experiment: six speakers are part of both groups.
d. Modified definite (experimental)

English prompt: “the red apples”

Target: \[ \text{DEF.PL} \ \text{red-DEF} \ \text{apple-DEF.PL} \]

Due to the nature of the experiment, some items were elicited several times in the same condition, since they refer to the protagonists of the story. Another consequence of the task is that the given responses sometimes deviate from the expected (target) response. These responses were included in the analysis, since the use of compositional definiteness can also be studied in NPs with a different lexical item than the targeted one.

3.2. Participants

Forty speakers participated in the translation task. Nineteen of them have been analysed for the current project. These were selected because they also participated in the picture naming experiment. The participants were born in the US between 1924 and 1946 as third or fourth generation immigrants. All are L1 speakers of Norwegian who shifted dominance to English from school age onwards. The participants come from the states Minnesota, Wisconsin and North Dakota. Four of them are women. In the remainder of the paper, examples are accompanied by a code referring to the speaker. The speaker code consists of the place name where the participant comes from, plus a unique number.

3.3. Analysis

All responses to the test items were transcribed and analysed as either target-like or non-target-like with respect to the definiteness marking. Gender marking was not taken into account in this analysis, which means, for example, that a definite noun containing the definite suffix was analysed as target-like, regardless of the gender of this suffix.

In a translation task, participants do not always translate as expected: elements are left out and pronouns are used. Responses containing pronouns or ellipsis of the noun, and complete English phrases were not analysed. The remaining responses were analysed with respect to the NP type used by the participant. The analysis of (non-)target-like thus only refers to the actual utterances of speakers and not to how accurately they translated the English prompt.

Simple and modified indefinite phrases were analysed as target-like when an indefinite article or an indefinite plural suffix was present. Simple definite phrases were analysed as target-like when the suffixed article was present, and as non-target-like when there was double definiteness or only a prenominal determiner. Modified definite phrases were analysed as target-like when both the prenominal determiner and the suffix were present. Cases with either one missing, or cases where the determiner was replaced by a demonstrative were analysed as non-target-like. The next section will present the results of this analysis.

4. Results

The individual results are shown in Figure 1, where for each participant the percentage of target-like utterances across the four conditions is given. As can be seen, all participants score very high percentages with respect to target-like utterances in the three control conditions (indefinite simple, definite simple, indefinite modified). Many speakers score 100% target-like in one or more control conditions. Except for

---

6 In the background interview, most participants indicated that they learned English when they went to school, which means that age of exposure to English is (approximately) 7. However, we cannot exclude the possibility that they have heard some English before they went to school.

7 Only complete English phrases were excluded from the analysis, phrases containing only one English word were analysed. Typically, these phrases contain Norwegian functional material (cf. Riksem 2017).

8 Some Norwegian nouns (mainly neuter ones) can have a bare indefinite plural: \textit{et hus} ‘a house’ – \textit{hus} ‘houses’. In these cases, a bare noun in indefinite plural condition was analysed as target-like.
one speaker, no one scores below 80% target-like in any of the control conditions. For simple definite NPs, all but one speaker score between 95-100% target-like. This indicates that these three conditions do not pose any problems for the speakers when it comes to definiteness marking.

However, on the modified definite structures the results are very different. No one scores 100% target-like and in fact, no one scores higher than 77%. Five speakers score more than 50% target-like, ten speakers score between 1 and 50% target-like and four people score 0% target-like. These speakers never produce modified definite phrases that are grammatical in homeland Norwegian.

Figure 1: Percentage target-like utterances. For each participant, target-like scores are given for all four conditions.

Table 1 summarizes these results on the group-level. It shows that the average score for the three control conditions is much higher than for the modified definite condition. Also, the number of participants with a 100% score and the range of scores differs much between the control conditions on the one hand and the experimental condition on the other hand. Note that the ranges of scores of control conditions and experimental condition do not overlap.

Table 1: Summary of the results. For each NP type, the average score, number of participants who score 100% target-like, range of scores and standard deviation are shown. Modified definite NPs are the test condition, the other three are control conditions.

<table>
<thead>
<tr>
<th>NP type</th>
<th>Average score</th>
<th>Standard deviation</th>
<th>n of participants with 100%</th>
<th>Range of scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple indefinite</td>
<td>97.6%</td>
<td>5.980</td>
<td>16</td>
<td>81.25 – 100%</td>
</tr>
<tr>
<td>Simple definite</td>
<td>98.0%</td>
<td>5.096</td>
<td>13</td>
<td>77.78 – 100%</td>
</tr>
<tr>
<td>Modified indefinite</td>
<td>93.1%</td>
<td>8.738</td>
<td>11</td>
<td>77.78 – 100%</td>
</tr>
<tr>
<td>Modified definite</td>
<td>38.7%</td>
<td>24.781</td>
<td>0</td>
<td>0 – 76.92%</td>
</tr>
</tbody>
</table>

The numbers in Table 1 clearly show that the phenomenon compositional definiteness is vulnerable to restructuring in AmNo. All speakers differ from homeland Norwegian, although there is individual variation in how many target-like utterances they produce. Fifteen of the speakers do use compositional
definiteness in some of their utterances. Two examples are given below\(^9\). The percentage behind the speaker code indicates the total percentage of target-like modified definite phrases of the speaker.

(3) a. de tre ung-ene (iola_WI_05gm, 76.92%)
    DEF.PL three child-DEF.PL
    ‘the three children’

b. den vesle jent-a (westby_WI_11gm, 50%)
    DEF.SG little.DEF girl-F.DEF.SG
    ‘the little girl’

In (3a), compositional definiteness is used in a phrase modified by a cardinal. This phrase is uttered by the speaker with the highest score of target-like utterances. In (3b), compositional definiteness is used in a phrase modified by an adjective.

As is clear from Figure 1 and Table 1, all speakers use non-target-like utterances to various degrees. Three different types of such non-target-like responses are given below.

(4) a. hvit-e hest-en (westby_WI_06gm, 0%)
    white-DEF horse-M.DEF.SG
    ‘the white horse’
    target: den hvite hasten

b. den små Jente (sunburg_MN_11gk, 71.43%)
    DEF.SG little Girl
    ‘the little girl’
    target: den små jenta

c. denne hvit-e hest-en (fargo_ND_01gm, 40%)
    DEM.SG white-DEF horse-M.DEF.SG
    ‘the white horse’
    target: den hvite hasten

The utterance in (4a) only contains the suffixed article, but the prenominal determiner is missing. The speaker quoted here does not use CD at all\(^10\), but the same structure is also found in speakers who use CD in other phrases. In (4b), the utterance does contain the prenominal determiner, but the suffixed article is missing\(^11\). Finally, in example (4c) a demonstrative instead of a determiner is used. This category is analysed as non-target-like, even though speakers could in theory intend to express a demonstrative meaning (in that case the utterance would be translated to ‘this white horse’)\(^12\).

Table 2 shows which types of structures were used by all participants during the experiment. In total, 41.1% of the utterances was target-like. This is comparable to the average score across participants (cf. Table 1). Besides the types illustrated in (3) and (4), two other types of target-like structures are used. These are adjective incorporations\(^13\) (very infrequent) and exceptions, in which the prenominal determiner is optional and often left out. Table 2 shows for each structure type the number and proportion of utterances belonging to this type, and how many participants use this structure at least once.

\(^9\) For the sake of simplicity, I present the examples in Bokmål (standard) Norwegian here. The data transcriptions however reflect the dialect spoken by the speaker, and the dialectal variation was taken into account when analysing a phrase as target-like or non-target-like.

\(^10\) In the spontaneous speech available in CANS, this speaker uses a modified definite NP with compositional definiteness. However, in the translation task he does not.

\(^11\) Besides that, the form of the adjective is unexpected (små ‘little’ is typically only used in plural), but that is not taken into account here.

\(^12\) As an anonymous reviewer suggested, the phrases containing a demonstrative could be excluded. This would not change the group results or the patterns described substantively. However, the individual results of two speakers that use the demonstrative often (3 times) would be different.

\(^13\) In these structures, an uninflected adjective is incorporated into a definite noun and the prenominal determiner is absent.
As can be seen in Table 2, compositional definiteness is used in less than a quarter of the modified definite phrases. The most frequently used structure is the one with a missing determiner (4a): it is used in 42.47% of the utterances. This category is larger than the total amount of target-like responses and is uttered at least once by 18 out of 19 participants. The mirror image, a phrase where the suffix is missing (4b), is much less frequent: it is used in 7.53% of the responses and by 6 participants. Five of them also use phrases without determiner in their other utterances. The overuse of demonstrative (4c) could be intentional. It is used in 8.9% of the utterances, and by eight speakers. Three of them use this structure more than once.

Table 2: Types of structures used in modified definite phrases. Total number, proportion, and how many participants use this type at least once is given. * = This structure is non-target-like.

<table>
<thead>
<tr>
<th>Type of Structure</th>
<th>n of utterances</th>
<th>Percentage</th>
<th>n of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compositional definiteness</td>
<td>34</td>
<td>23.29%</td>
<td>14</td>
</tr>
<tr>
<td>Adjective Incorporation</td>
<td>1</td>
<td>0.68%</td>
<td>1</td>
</tr>
<tr>
<td>Exception</td>
<td>25</td>
<td>17.12%</td>
<td>13</td>
</tr>
<tr>
<td>* Missing determiner</td>
<td>62</td>
<td>42.47%</td>
<td>18</td>
</tr>
<tr>
<td>* Missing suffix</td>
<td>11</td>
<td>7.53%</td>
<td>6</td>
</tr>
<tr>
<td>* Overuse of demonstrative</td>
<td>13</td>
<td>8.90%</td>
<td>8</td>
</tr>
</tbody>
</table>

The results make clear that CD is vulnerable to restructuring to some extent in all speakers. Phrases without the determiner are the most frequent type of structure. For some speakers, phrases without the suffix are used as an alternative.

5. Conclusion and discussion

The first goal of this paper was to investigate CD in an experiment, to complement the studies on spontaneous speech. The results clearly show that while the speakers score very much target-like on noun phrases not requiring CD, they all score much less target-like on phrases requiring CD (modified definite phrases). There is large variation among the speakers, but CD is vulnerable in all of them. This corroborates the findings from Johannessen (2015b) and Anderssen, Lundquist & Westergard (2018/fc.) in spontaneous speech.

The second aim of this paper was to find which element in CD is most vulnerable in the speakers. I found that utterances without the prenominal determiner were most frequent and used by virtually all speakers. Utterances without the suffixed article were much less frequent and used as an alternative by some speakers. This finding is interesting when we take into account the dominant language of all speakers. In English, definiteness is always marked by a prenominal determiner and there is no suffixed article. The fact that this ‘English-like’ structure is not very frequent and only used by some participants means that the vulnerability of CD is not caused by transfer. Even though English was very present during the experiment, English-like phrases were not frequent at all.

This result is unexpected in the light of Riksem’s (2017) findings. Riksem finds a growth in use of phrases without a definite suffix between older stages of AmNo (Haugen 1953) and current speakers (CANS). This difference in results might be explained by the fact that Riksem (2017) studies mixed phrases with English content words, whereas the present paper studies Norwegian phrases. Further research will have to show whether this explanation is on the right track.

Seen in the light of cross-linguistic overcorrection (Kupisch 2014), this result is less unexpected. In her study on adjective placement in German-Italian bilinguals, Kupisch (2014) found that the heritage speakers of Italian tended to overuse the order N-Adj, which is the structure that is different between the two languages. As a result of this overcorrection, the heritage speakers produce phrases that are divergent from the homeland variety but also divergent from the dominant language.
Something similar is found in the current study: the speakers of American Norwegian do not use CD in modified definite phrases (i.e. they diverge from homeland Norwegian), but do not use English-like phrases either. Instead of omitting the definiteness morpheme that is different between the two languages (the suffixed article), they omit the morpheme that is found in both languages (the determiner). It has to be noted that the speakers in Kupisch (2014) are not elderly speakers of a moribund language variety, and therefore quite different from the speakers in my study. They could however still be influenced by cross-linguistic overcorrection, as is also found for possessives (Anderssen, Lundquist & Westergaard 2018/fc.).

The participants in the present study seem to fall in two groups: one that uses structures without the prenominal determiner and one that also uses structures without the suffixed article. This finding is in line with the two groups found in Anderssen, Lundquist & Westergaard (2018/fc.), who argue that the first group is influenced by cross-linguistic overcorrection whereas the second group is influenced by cross-linguistic influence (transfer). Besides cross-linguistic overcorrection, frequency also seems to play a role. The suffixed article is more frequent in Norwegian than the prenominal determiner, so the speakers who produce modified definite phrases without the determiner are not only using the non-English structure but also the most frequent definiteness marker.

Anderssen, Lundquist & Westergaard (2018/fc.) furthermore suggest that the two groups of speakers exemplify two different proficiency groups. The group that is influenced by cross-linguistic overcorrection is argued to be more proficient, whereas the group that is influenced by transfer is less proficient. However, proficiency has not been measured for these speakers. Anderssen, Lundquist & Westergaard (2018/fc.) therefore use the amount of phrases with CD and use of target-like gender marking as measures of proficiency. With respect to these two factors, the differences between the groups are not statistically significant, although the error counts suggest that there is a small proficiency difference between the groups. Further research, preferably with independent proficiency measurements, is needed to draw conclusions about the influence of proficiency. Nevertheless, it is clear that only a small group of the speakers is influenced by transfer from English and that most speakers could be influenced by cross-linguistic overcorrection.

The data presented here show the vulnerability of compositional definiteness in American Norwegian. Some speakers utter phrases that can be explained by transfer, but the typical AmNo modified definite phrase is one without the prenominal determiner, which cannot be the result of transfer. Further research into what exactly causes the vulnerability of CD and the factors of influence is work in progress.

References


Kupisch, Tanja. 2014. Adjective placement in simultaneous bilinguals (German-Italian) and the concept of cross-linguistic overcorrection. *Bilingualism: Language and Cognition* 17(1). 222-233.


