Variation between the infinitival complementizers *om/voor* in spontaneous speech data compared to elicitation data

Leonie Cornips

Meertens Instituut

1 ▪ INTRODUCTION

In several previous dialect atlas projects in The Netherlands and in the Dutch-speaking part of Belgium, the focus has always been on lexical, phonological and morphological variables.[1] This corresponds to atlas projects all over the world in which no more than five percent of published dialect maps involve syntactic data (Gerritsen 1993:343). There are two well-known reasons why syntactic research is always neglected in dialect atlas projects or in dialect research more generally (cf. Bucheli and Glaser in this volume). The first one has to do with the views of more traditional dialectologists that syntactic variation in the Dutch dialects is barely visible and, subsequently, is negligible by comparison to lexical, phonological and morphological variation. Apart from atlas projects, which are a more traditional dialect enterprise, it is only a very recent development in the Netherlands and in Belgium that generative linguists take the dialect as an object of research. The generative framework differs from that of dialectology in that the former considers syntactic features always in interaction with other variants as predicted by theory whereas the latter examines syntactic variants in isolation. The second reason is linked to the fact that special methodology is required to obtain syntactic dialect data from a large geographical area regardless or whether this data is needed to device a linguistic atlas or as a contribution to theory refinement. The variationist approach diverges from both traditional dialectology and generative approaches since it focuses on achieving representativeness regarding the speakers observed; obtaining controlled recordings of vernacular speech and collecting a substantial quantity of data in order to achieve descriptive and observational adequacy (cf. Cornips and Corrigan t.a.).

.....

[1] I like to thank Karen Corrigan for her valuable suggestions and correction of the English of the first three sections of this paper.

However, it is nearly impossible to collect data from a large geographical area using these methods. Apart from the amount of time required to collect spontaneous speech data that is truly vernacular across a large geographical area, there are other, often insurmountable problems, namely (i) syntactic tokens are infrequent and they do not involve the whole range of possible variants, that is to say, the syntactic tokens do not exhibit the complete paradigm, (ii) the syntactic variables do not always show up in interaction with other relevant syntactic variables that are predicted by theory and (iii) negative data are lacking. As a result, in addition to systematic recording of spontaneous speech and introspective judgements, an approach is needed in which dialect data are collected by other elicitation techniques such as written questionnaires and oral tasks in order to achieve greater observational and explanatory force.

In this paper, I will discuss the discrepancies between spontaneous speech data and elicitation data with respect to the linguistic and social distribution of syntactic variants. The structure of this paper is as follows: Section 2 discusses the unreliability of introspective judgements; Section 3 deals with the task-effects of written questionnaires that are used to collect syntactic dialect data in a large geographical area such as the written questionnaires used in the SAND-project (syntactic atlas of Dutch dialects, cf. Cornips & Jongenburger 2001). In the last section, I will discuss the discrepancies between spontaneous speech and oral elicitation data with respect to the linguistic and social distribution of linguistic variables, focussing on the variation between the infinitival complementizers *om* and *voor* in Heerlen Dutch. These empirical data concerning the complementizers *om* and *voor* shed more light on the answers to the questions whether the variation between *om* and *voor* is a syntactic and/or a semantic phenomenon or just a phenomenon of lexical substitution.

2 ▪ UNRELIABILITY OF INTROSPECTIVE JUDGEMENTS

Generativists focus on native-speaker introspection in an idealised environment in their pursuit of explanatory adequacy. Direct questions about the (un)grammaticality of syntactic features may provide insight into a speaker's competence far more readily than spontaneous speech data do. However, since Labov (1972), the unreliability of native speaker judgements is well known. He has shown that: "*whenever a subordinate dialect (stigmatized) is in contact with a superordinate dialect (prestige), answers given in any*

formal test situation will shift from the subordinate towards the superordinate in an irregular and unsystematic manner (1972: 21)". Moreover, "*speakers' attitudes towards well-established linguistic variables will also be shown in self-evaluation tests. When asked which of several forms is characteristic of their own speech, their answers reflect the form which they believe has prestige or is "correct" rather than the form they actually use. ...this kind of test data cannot be interpreted without data on the subjects' actual speech patterns*" (Labov 1972: 213). Thus, one of the conditions that promote the failure of linguistic intuitions is social intervention, that is to say, when a socially superordinate norm takes precedence over the native system (Labov 1996: 100). These findings are particularly relevant in a setting in which two or more language varieties or dialects differ in social prestige as in the English vernacular observed in the United States. With respect to the European dialect context, the Dutch case represents the same language situation, that is: a diaglossia type in which the standard language is both spoken and written showing vertical and horizontal levelling in the most widespread context (cf. Auer 2000). It is obvious that in the Dutch context, especially in the western part of the Netherlands in which there is one, single style continuum between the dialect and the standard language, the social intervention condition on the reliability of elicited speaker judgements may play an important role. Hence, there is no sharp discontinuity between the local vernacular and the standard language (cf. Hinskens 1996: 132). However, this condition may not be relevant in a medial diglossia situation in which the dialect is spoken and the standard language is exclusively a written variety, such as Nynorsk in Norway and Swiss standard German. These dialects have high prestige and regarding code-switching, they are equal partners and show a balanced repertoire (cf. Auer 2000). Clearly, spontaneous speech data and elicitation data such as introspective judgments may differ due to the social prestige of the dialect varieties in relation to the standard language.

3 ▪ THE TASK-EFFECTS OF WRITTEN QUESTIONNAIRES

An important experimental method in order to construct a dialect atlas or to collect dialect data in a large geographical area is the use of written questionnaires. In each location relevant to the research, a single native speaker, that is to say, a so-called NORM (**N**on-mobile, **O**lder, **R**ural, **M**ale speaker) for whom the local dialect is native

is asked to complete a questionnaire. This method has the advantage of systematically gathering dialect data in a large geographical area within a short time span. Moreover, it is an elicitation technique that enables the researcher to standardize both the collection of, and the analysis of, the material. However, this method induces numerous well known task effects, which are briefly discussed below (cf. Greenbaum 1973, Oxford 1982).

The first task effect is immediately obvious: that is, in general, standard languages are written varieties whereas dialects are spoken ones. Subsequently, the written response of the informant will often be unduly influenced by prescriptive educational practices.

Secondly, most written questionnaires contain a translation task whereby the speaker has to 'translate' constructions from the standard language into the local dialect. One task-effect, which this action is inclined to induce is the repetition-effect; that is, the standard construction will be translated literally into the local dialect. Hence, the speaker has to write in dialect, although he is not used to doing so. As a result, s/he will concentrate more on the task of spelling and translating dialectal lexical elements not conventionally represented in the standard language with the result that more of their attention is focused on completing this insignificant aspect of the task from the researcher's perspective who ia actually more interested in the resp;ondent's ability to handle syntactic variation.

The third task-effect is that grammatical constructions on the syntactic level may be judged to be ungrammatical simply on the basis of lexical items. Figure 1 illustrates this task-effect. In 1995, a questionnaire was sent out by the Meertens Instituut in the Rhineland area and it was offered in standard German. For each location, one native speaker of the local dialect filled in the questionnaire. In this Rhineland questionnaire both the impersonal and adjunct middle with and without the reflexive *sich* were administered (cf. Cornips 1996b, Cornips & Corrigan 2001). For each variant (a), (b), (c), and (d) in figure 1, the native speakers were asked to answer the following two questions. The first question was: do you ever 'encounter' the variant in your local dialect 'kommt vor/ ist manchmal zu hören'? Furthermore, the speakers were asked if they considered the variant to be the most 'common' one in their local dialect 'am gebräuchlichsten'. In addition, the native speaker was asked to give a translation 'Übersetzung' of the most common construction in his dialect.

FIGURE 1 ▪ *Part of a written questionnaire based on syntactic variants e.g. the middle constructions*

		1		2
		kommt vor / ist manchmal zu hören		am gebräuchlichsten
a	Dieser Stuhl sitzt herrlich	~~ja~~	nein	a
b	Dieser Stuhl sitzt sich herrlich	~~ja~~	nein	b
c	Es sitzt sich herrlich auf diesem Stuhl	~~ja~~	nein	c
d	Es sitzt herrlich auf diesem Stuhl	~~ja~~	nein	d
3	Übersetzung …….: *"Herrlich" = total ungebräuchlich bezw. Nur aus der Bibel bekannt (Übersetzung nach Hermanns = briljant.)*			

The comments of the speaker in the *Übersetzung* phase in figure 1 reveals that he doesn't consider any of the middle constructions to be acceptable due to the fact that the lexical item *herrlich* 'pleasantly' is unknown in the local dialect 'total ungebräuchlich'. However, the reflexive impersonal construction in (c) is fully grammatical in the Rhineland dialect and in standard German.

The fact that the native speaker judges a certain form to be completely unacceptable but can nevertheless, be recorded using freely in every-day conversation, is a striking task-effect of both elicited introspective judgements and written questionnaires which inquire about the (relative) acceptability of a given construction (Labov 1996: 78). Thus, the translation task of the most common sentence, in addition to questions about acceptability, shed light on the issue of why the native speaker judges a specific construction to be unacceptable. Moreover, it provides more insight into the reason why the native speaker is understanding or interpreting a specific construction, in a specific way which was not intended by the researcher, as illustrated in figure 2. The same Meertens questionnaire in the Rhineland area reveals that only one speaker (out of nineteen) judged the argument middle in the perfect without a reflexive in (c) as acceptable ('encounter' and 'most common' see bold print) in the dialect (cf. Cornips and Corrigan t.a.):

FIGURE 2 ▪ *Part of a written questionnaire based on syntactic variants e.g. the middle constructions*

		1 kommt vor/ ist manchmal zu hören	2 am gebräuchlichsten	
a	Das Bier hat *sich* gut getrunken gestern abend	ja	nein	a
b	Das Bier trank *sich* gut gestern abend	ja	nein	b
c	Das Bier hat gestern abend gut getrunken	**ja**	nein	**c**
3	Übersetzung: Dat Bier hat sich jestere Ovend jot drenke losse.ov.: dat Bier hat jestere Ovend jot jeschmaht			

However, the speakers' translations (Übersetzung) demonstrate that he has interpreted this middle as a regular reflexive *lassen* ('let')-construction and/or as a totally different construction.

Finally, a specific task-effect is an order-effect. The more the speaker has to make judgements about a specific construction that is offered several times in a different way in one questionnaire, the more he will judge it to be acceptable. An alternative to this task-effect is to vary the order of the questions per list (cf. Bree 1981, Cornips 1994: 46). Note, that this is only an alternative if there are more informants present at each location. Otherwise, the variation due to the different orders of the questionnaires may be attributed to the geographical spread of the questionnaires. themselves

In conclusion, written questionnaires have different task-effects that must be considered in any resultant analysis.

4 ▪ THE DISCREPANCIES BETWEEN SPONTANEOUS SPEECH DATA AND ORAL ELICITATION DATA

In this section, I will discuss the discrepancies between spontaneous speech data and oral elicitation data (Bock 1986). To this end, I will present a subpart of a study on the social dimensions of the regional Dutch spoken in Heerlen (henceforth: Heerlen Dutch), focusing on the variation between the infinitival complementizers *om* and *voor* (cf. Cornips 1994, 1996a). Heerlen is situated in the province of Limburg in the southeast of The Netherlands, near the Belgian and German border and it is the result of a process of abrupt language shift in the beginning of this century with the local dialect as the source and standard Dutch as the target language.

4.1 ▪ Spontaneous speech data

The data consist of 33,5 recorded hours of spontaneous speech between two speakers who did not know each other but they belonged to the same cell (in-group conversation) and the recordings took place at the speakers' homes. A total of 67 male speakers participated in this survey.[2] Later, I will come back in more detail to the sociolinguistic part of this study.

In Heerlen Dutch, two variants of the infinitival complementizer arise, *om* is the standard Dutch variant realization, while *voor* is the local dialect variant realization (Jongeneel 1884: 65, Cornips 1994).[3] The data reveal variation to a large extent. As can be seen in figure 3, some speakers produce only *om* (for example, speaker 1 and 2 in the first column), some speakers use both *om* and *voor* (for example, speaker 4 in the first column) and one speaker uses only *voor* (speaker 67 in the first column):[4]

.....

[2] Only one informant had a conversation with the interviewer.

[3] Only the data showing an overt complementizer *om* or *voor* are taken into consideration.

FIGURE 3 ▪ *Individual production of* om *and* voor

Individual Male Speaker	no. *OM*	no. *VOOR*	Total	proportion *VOOR*
1- 2	19	0	19	0
3	19	5	24	.21
4	19	1	20	.05
5	18	0	18	0
6	17	0	17	0
7	16	0	16	0
8	16	3	19	.16
9	15	1	16	.06
10	14	0	14	0
11	14	4	18	.22
12	14	1	15	.07
13-14	13	0	13	0
15	13	1	14	.07
16	12	0	12	0
17	12	1	13	.08
18-19	11	0	11	0
20	11	2	13	.15
21	11	1	12	.08
22	11	3	14	.21
23	10	0	10	0
24-28	9	0	9	0
29	9	5	14	.36
30	9	1	10	.10
31-32	8	1	9	.11
33	7	2	9	.22
34-35	6	0	6	0
36	6	3	9	.33
37	6	1	7	.14
38-40	5	0	5	0
41-42	5	5	10	.50
43-45	4	0	4	0
46	4	2	6	.33
47	4	1	5	.20
48	4	10	14	.71
49-52	3	0	3	0
53	3	2	5	.40
54-55	3	1	4	.25
56	2	6	8	.75
57-58	2	4	6	.67
59	2	2	4	.50
60	1	0	1	0
61	1	8	9	.89
62	1	7	8	.88
63	1	5	6	.83
64	1	2	3	.67
65-66	1	1	2	.50
67	0	16	16	1
Total	523	120	643	

Let us examine in more detail the syntactic factors that might affect the occurrence of *om* or *voor* in Heerlen Dutch as presented in figure 3.

[4] All speakers are male (see also 4.2).

First, both *om* and *voor* arise if the infinitival clause is an object clause, as in (1). The infinitival clause is the complement of the verb *proberen* 'try':

(1) a ik heb wel tachtig keer geprobeerd *om* met
I have ADV *eighty times tried* COMP *with*
dat roken te stoppen (15: Jan)[5]
that smoking to stop
'I have tried to quit smoking eighty times'

b. hebben een paar keer geprobeerd *voor* mee te
have a few time tried COMP PART *to*
doen aan zo'n quiz (2: Wybe)
do in such a quiz
'several times (we) tried to participate in such a quiz'

Both *om* and *voor* show up if the infinitival clause is used as an adjunct. In (2), the infinitival clauses are purpose clauses:

(2) a. je ziet vaak een taxi van Aken hier in
you see often a cab of Aken here in
Heerlen rondrijen ...ja *om* het spul op te
Heerlen rounddriving yes COMP *the stuff* PART *to*
halen
collect
'in Heerlen you often see a cab from Aken driving around to collect the

b. die moest naar de hei gaan elke dag *voor* te
he had PART *the heath go every day* COMP *to*
wandelen *voor* lucht te krijgen (14: Ralph)
walk COMP *air to get*
'every day he had to go to the heath to walk, to get air'

[5] The figure appearing before the speaker's pseudonym refers to the number of the tape-recording.

Further, both *om* and *voor* appear if the infinitival clause is the complement of an adjective, such as *moeilijk* 'difficult' in (3a):

(3) a. is ( )moeilijk *om* Sinterklazen te vinden (10: Nico)
is difficult COMP *Santa Clauses to find*
'is difficult to find Santa Clauses'

b. ik vind het zo erg *voor* een beest
I find it so terrible COMP *an animal*
kapot te maken (31: dhr Bast)
RESULT *to make*
'I find it so terrible to destroy an animal'

In addition, both *om* and *voor* are used with verbs of object control and arbitrary control, as illustrated in (4) and (5), respectively:

(4) a. dan riep ik m'n vrouw naar .. het raam *om* te
then called I my wife to the window COMP *to*
kijken als d'r een fietser langskwam (7: Ruiter)
look if there a cyclist came along

b. en dan wil ik je nou de kans geven
and then want I you adv the opportunity give
voor nog wat bij te verdienen (4: dhr Mije)
COMP ADV *something to earn*
'I want you to give the opportunity to earn an additional income'

(5) a. dat blijft heel moeilijk *om* d'r in te geloven (21: Mart)
'It remains very difficult to believe in

b. het is nou tijd he *voor* te stekken (2: Wybe)
it is ADV *time uh* COMP *to strike cuttings of plants*

In addition, both *om* and *voor* show up in infinitival relative clauses:

(6) a. . . leuke jongen *om* mee op te schieten maar hij kan geen nee hebben (5: André)
'nice boy to go along with but he can't stand a 'no'-answer'

b. die groep van eh vijftien tot achttienjarigen (…) .. dat is best wel aardig *voor* mee te werken ja dus eh (19: Cor)
'this group of youngsters aged between fifteen and eighteen is nice to work with'

Strikingly, the syntactic variation between *om* and *voor* manifests itself at the individual level even while one speaker is maintaining the same level of speech style (Bickerton 1971 (cf. Figure 3). Consider the constructions in (7) that were uttered by one speaker ('Bert') on the same occasion. It appears that the same speaker uses both *om* and *voor* in the same linguistic category: that is, the verbs *(terug)komen* in both (7a) and (7b) are verbs of subject control and both sentences are infinitival purpose clauses:

(7) a. moet ik terug komen *om* dat half jaar ( ) af te
must I back come COMP *that half year* PART *to*
maken (12: Bert)
finish
'I must come back in order to finish that half year'

b. je komt hier *voor* te studeren (12: Bert)
you come here COMP *to study*
'You will come here in order to study'

From the above, it will be clear that in Heerlen Dutch both at the individual and at the group level, *voor* has the same syntactic distribution as *om*. The similar syntactic distribution of *om* and *voor* suggest that they are only lexical variants.

However, it is worthwhile to find out certain tendencies selecting the complementizer *voor* or *om*. First, let us consider standard Dutch. In the spoken variety of standard Dutch, in addition to the infinitival complementizer *om*, *voor* is also marginally used. It has been pointed out in the literature that *voor* in Dutch dialects is sometimes used if the infinitival clause is a purpose clause as in (3) (Gerritsen 1991: 61, 69). Since in Dutch dialects (as is the case in standard Dutch) the preposition *voor* is also capable of expressing purpose (Geerts 1984: 881), it can be argued that the form *voor* in a purpose infinitival clause corresponds semantically to the form *voor* used as a preposition. Therefore, it may be argued that in Dutch there exists a syntactic/semantic factor that promotes or inhibits the occurrence of the complementizer *voor*. Let us now consider whether this factor is also relevant in

Heerlen Dutch with regard to the occurrence of the complementizer *voor*. Table 1 reveals the data of speakers who use both *om* and *voor* occurring in the two types of infinitival clauses, namely [+/- purp] infinitival clause. It displays that in spontaneous speech the factor [+purp] yields significant results with the use of the dialect variant *voor*: that is, *voor* is used more frequently in a [+purp] clause.

TABLE 1 ▪ *The distribution of* om *en* voor *according to [+purp] and [-purp]-infinitival clause (spontaneous speech)*

	OM	VOOR	total
- purp	149	21	170
+ purp	104	99	203
	253	120	373

$x^2(\pm$ purp) = 56.21 df=1 $p<.001$

The range of individual grammars in Heerlen Dutch based on 643 *om/voor* occurrences in spontaneous speech, 67 speakers and 2 infinitival-types, e.g. [+purp] and [-purp] clauses reveals that the speakers produce eight different grammars that take the form *om* or *voor* or *om/voor* in each linguistic category (see also Bickerton 1971). Furthermore, 28 speakers produce only *om*, whereas 22 speakers produce *om* in a [-purp] clause but *om/voor* in a [+purp] clause, respectively grammar 1 and 2. Note that the speakers do not produce grammar 9. This behaviour confirms, to a certain extent, the results shown above; if both *om* and *voor* are simultaneously available, an infinitival non-purpose clause promotes the variant *om* and an infinitival purpose clause promotes the variant *voor*.[6]

[6] Only grammar 8 in figure 4 is a counterexample.

FIGURE 4 ▪ *Number of individual grammars according to linguistic category on the basis of spontaneous speech*

Grammar	*[-purp]* clause	*[+purp]* clause	number of speakers (N=67)
1	om	om	28
2	om	*om/voor*	22
3	om	voor	3
4	*om/voor*	voor	3
5	*om/voor*	*om/voor*	3
6	voor	*om/voor*	3
7	voor	voor	1
8	*om/voor*	om	1
9*	*voor	*om	0
	-----	*om/voor*	1
	-----	om	1

Five out of eight grammars display inherent variability, namely the grammars 2, 4, 5, 6, 8. I argue that the inherent variability in Heerlen Dutch is due to a semantically controlled dialectization of the spoken variety of standard Dutch. In other words, influenced by the use of *voor* both in the local dialect and in the spoken variety of standard Dutch, the speakers do not derive the variant *om* but the variant *om/voor* from standard Dutch. Since in standard Dutch the variant *om/voor* is most frequently used in an infinitival purpose clause, only this option explains clearly the large number of realisations of grammar 2.

4.2 ▪ The social stratification of voor-usage in spontaneous speech.

In standard Dutch the use of *voor* in an infinitival purpose clause is due to a recent language change. From the above, it is not obvious in Heerlen Dutch whether, from the synchronic point of view, the use of *voor* in an infinitival purpose clause originates from the local dialect or from a recent development in standard Dutch.

In contrast, it is certain that the use of *voor* in a non-purpose clause originates exclusively from the local dialect since this use is not acceptable in standard Dutch, that is to say, it is rejected by prescriptive norms. With the above assumptions in mind, we expect that (i) the group of dialect speakers will produce more *voor* in a non-

purpose clause than the other groups of speakers and (ii) *voor* in a purpose clause will be more frequently used by the younger speakers.

By means of the sociolinguistic part of this study, these hypotheses can be confirmed or rejected. In this survey, the speakers were divided into three language groups according to their language background, namely immigrant (N=19), dialect (N=29) and Heerlen Dutch (N=19). Immigrant speakers spoke (Heerlen) Dutch as a first language, and their parents were born outside the province of Limburg. Dialect speakers spoke the local dialect as a first language and (Heerlen) Dutch as a second language. Heerlen Dutch speakers spoke (Heerlen) Dutch as a first language and, their parents spoke the local dialect as a first language. The speakers were then further subdivided into smaller groups according to their education/occupation and age. The education/occupation variable is based on two values on a high to low scale, i.e. middle/high level employees (N= 39) and unskilled/skilled labour (N=28). With respect to the age, 'young' speakers (aged between 20 and 45, N= 37)) were distinguished from 'old' speakers (aged over 60, N= 30). The speaker variables are shown in table 2. The specification of these variables made it possible to investigate whether the groups of speakers show any social stratification.

TABLE 2 ▪ *Number of speakers by speaker variables*

	low level of education		*high* level of education		total
	young	*old*	*young*	*old*	
language					
immigrant	3	6	5	5	19
dialect	5	6	8	10	29
Heerlen Dutch	8	--	8	3	19
total	16	12	21	18	N=67

The quantitative analysis confirms the hypotheses. 73% of the speakers producing *voor* in a non-purpose clause belong to the group of dialect speakers and 82% belong to the group of older speakers. Sure enough, this distribution supports the hypothesis that the use of *voor* in a non-purpose clause originates from the local dialect. Moreover, as expected, 62% of the speakers producing *voor* in a purpose infinitival clause belong to the youngest age group (cf. Cornips 1996a).

4.3 ▪ Oral elicitation data

The data in this sociolinguistic survey were also collected by means of a simple oral repetition test. In the repetition test, I offered 66 speakers 11 infinitival clauses.[7] Five of the infinitival sentences were purpose clauses and six non-purpose clauses (cf. Cornips 1994, 1996, see appendix). All of the speakers were asked to repeat sentences that contained more than twenty items or words containing either the *om* or *voor* complementizer. It can be argued that the large number of items has an effect on the capacity of the speakers' short-term memory. It is reasonable to assume that this kind of stimulus makes it difficult to repeat exactly the structure and, as a result, the speaker has to rely on his own grammar. If this is so, an accurate repetition of *om* or *voor* provides weak evidence that the test variant is within the speaker's dialect; consistent inaccuracies (usually translation, e.g. substitution of *om* for *voor* or the reverse) provide strong evidence that the test variant, e.g. *om* or *voor*, is not within his dialect (Carden 1976: 101).[8]

It is evident from table 3 that in the test data the factor [± purp] clause yields significant results with the dialect variant *voor*. As in the spontaneous speech data, *voor* is used more often in a [+purp] clause than in a [-purp] clause, namely 20% and 6% respectively.

TABLE 3 ▪ *The distribution of* voor *in test data by [-purp] and [+purp] infinitival clause (the* numerator *includes repetitions of* voor *and translations of* om *by* voor *whereas the* denominator *includes the total output of* om *and* voor

	[- purp] clause	*[+ purp] clause*
repetition *voor*	19/168 (11%)	39/169 (23%)
translation *om* by *voor*	2/160 (1%)	33/198 (17%)
total	21/328 (6%)	72/367 (20%)

x^2 (±purp) = 26.10, df = 1, p<.001

.....

[7] One older informant had many difficulties to hear the test sentences. This informant is not involved in the procedure of the elicitation tests.

[8] One reviewer pointed out to me that "neurologists use retarded repetition which has to be produced at least five seconds after the stimulus. There is evidence that after this period the speaker cannot use his memory and has to 'go through' his own linguistic knowledge, even in the case of very simple sentences." Of course, this type of test might be used here also.

However, no significant correlations are found with respect to the social distribution of the complementizer *voor*, and as a result, the social distribution of *om*/*voor* in the testdata differs from the spontaneous data. I argue that these different patterns are primarily due to distinct test effects. In order to provide some insight into these test effects, I present a detailed list of all the speakers with their variables who have a deviant output in the test data compared with their spontaneous speech in figure 5.

It is clear from figure 5 that a total of 25 speakers (one third of the speakers) display a different test output in comparison to their spontaneous speech data. There are two ways in which the speakers may vary in their test production. First, they use only *om* in their spontaneous speech whereas their test data display also *voor* (see group I, II and III in figure 5) or they realise only *voor* in spontaneous speech while they use only *om* in the test (see group IV). With respect to the seven speakers in group I and II, I argue that their translations of *om* by *voor* provide strong evidence that the infinitival complementizer *voor* is within their language variety, although it is not present in their spontaneous speech. Of course, the fact that the variant *voor* does not arise in their spontaneous speech does not imply that it does not belong to the grammar of these speakers "since nonoccurrence in a corpus may always be due to nongrammatical, contextual factors or even to chance" (Kroch 1989:200). Further, the output of the eighteen speakers in groups III and IV displays interesting test effects. Consider first group III. Since these six speakers only repeat the test input but do not translate it, their test production can be ascribed to a repetition effect (Bock 1986).

If we examine the speaker variables in this group, we see that proportionally, the majority of these six speakers belongs to the oldest age group (OLD: 4/29 = .14 versus YOUNG: 2/38 = .05), to the immigrant speakers (IMM: 4/29 = .21, HD = 1/19 = .05, DIA= 1/28 = .04) and to the group of speakers with a high level of education/occupation (HIGH: 4/38 = .11 versus LOW 2/28 = .07).

FIGURE 5 ▪ *Number of speakers according to variables who display a deviant test output in comparison with their spontaneous speech (N=25)*

		spontaneous speech	repetition *voor*	translation of *om* by *voor*	N	
I	IMM-young-low	*om*	yes	yes	1	
	DIA-young-high	*om*	yes	yes	1	N=2
II	IMM-young-high	*om*	no	yes	1	
	IMM-old-high	*om*	no	yes	1	
	IMM-old-low	*om*	no	yes	1	
	HD-old-high	*om*	no	yes	1	
	HD-young-high	*om*	no	yes	1	N=5
III	IMM-young-high	*om*	yes	no	1	
	IMM-old-low	*om*	yes	no	1	
	IMM-old-high	*om*	yes	no	2	
	DIA-old-high	*om*	yes	no	1	
	HD-young-low	*om*	yes	no	1	N=6
IV	IMM-old-high	*voor*	no	no	1	
	IMM-young-high	*voor*	no	no	1	
	DIA-old-high	*voor*	no	no	5	
	DIA-old-low	*voor*	no	no	1	
	DIA-young-high	*voor*	no	no	2	
	HD-young-high	*voor*	no	no	1	
	HD-young-low	*voor*	no	no	1	N=12

Group IV displays another performance-mechanism since these twelve speakers neither repeat nor translate the test input. As a result they use only *voor* in their spontaneous speech but the variant *voor* does not show up in their test output. We can conclude from this that these speakers are subject to stylistic variation. Their test output shows that they are capable of controlling variation between *om* and *voor.* These speakers are subject to the social intervention condition since *om* is the prestige variant belonging to the domain of standard Dutch (see §2). The majority of the speakers who are able to control the stylistic variation belong to the group of speakers with a high level of education/occupation (HIGH: 10/38 = .26 versus LOW: 2/28 = .07),

to the older speakers (OLD: 7/29 =.24 versus YOUNG 5/37 = .14) and the dialect speakers (DIA: 8/28 = .29, IMM/HD: 2/19 = .11).

Another striking result is that proportionally there are hardly any Heerlen Dutch, e.g. five out nineteen speakers (.26) who display a different test output in comparison to their spontaneous speech. From this, we may conclude that these five speakers use both *om* and *voor* to a great extent since the speakers in figure 5 use only *om* or *voor* but not both in their spontaneous speech.

5 ▪ CONCLUSION

The infinitival complementizers *om* and *voor* have the same syntactic distribution as *om* in Heerlen Dutch at the individual and the group level. Both variants appear with verbs of subject, object and arbitrary control. Further, *om* and *voor* arise in adjunct and complement infinitival clauses. The similar syntactic distribution of *om* and *voor* suggest at first sight that they are only lexical variants. However, the sociolinguistic part of this study shows that if both *om* and *voor* are simultaneously available, an infinitival non-purpose clause promotes the variant *om* and an infinitival purpose clause promotes the variant *voor*.

Moreover, it is shown that there are no discrepancies between spontaneous speech and oral elicitation data with respect to the linguistic distribution of the infinitival complementizers *om* and *voor*. However, the social distribution of the test output differs from the social distribution of the variants in spontaneous speech. I have argued that these differences are primarily due to different test effects. What is more, it appears that speakers with a high level of education and occupation in particular are capable in controlling their test output in contrast to their spontaneous speech. Apparently, the method of elicitation - in this case study a repetition test - determines the social stratification of the speakers, and hence, it is worthwhile to take this effect into account in examining syntactic variation.

Auer, Peter (2000). A European perspective on social dialectology. Talk presented at *First International Conference on Language Variation in Europe (ICLaVE1*). July, 1 Barcelona.

Bickerton, Derek (1971). Inherent variability and variable rules. *Foundations of language* 7. 457-92.

Bock, J. Kathryn (1986). Syntactic persistence in language production. *Cognitive Psychology* 18: 335-387.

Bree, C. van (1981)**.** *Hebben-constructies en datiefconstructies binnen het Nederlandse taalgebied: een taalgeografisch onderzoek*. Leiden: Dissertatie

Bucheli, C. and E. Glaser. (This volume). The syntactic atlas of Swiss German dialects: empirical and methodological problems.

Carden, Guy (1976). Syntactic and semantic data: replication results. *Language in Society* 5. 99-104.

Cheshire, J. (1987). Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25, 257-282.

Cornips, Leonie (1994). Syntactische variatie in het Algemeen Nederlands van Heerlen (Syntactic variation in Heerlen Dutch). Amsterdam: Dissertation, IFOTT 6.

¾¾ (1996a). Social stratification, linguistic constraints and inherent variability in Heerlen Dutch: The use of the complementizers om/voor. In: Arnold, J., R. Blake, B. Davidson, S. Schwenter & J. Solomon (eds.), *Sociolinguistic Variation: Data, Theory, and Analysis*. Selected papers from NWAVE-23, CSLI Publications: Stanford University. 453-467.

¾¾ (1996b). The spread of the reflexive adjunct middle in the Limburg dialects: 1885 -1994. *Linguistics in the Netherlands 1996* (Cremers C. & M. Den Dikken eds.) 49 -60.

Cornips, L. & K. Corrigan (to appear). Convergence and divergence in grammar. In: *The convergence and divergence of dialects in contemporary societies* (P. Auer, F. Hinskens & P. Kerswill eds.). Cambridge University Press.

Cornips, L. & W. Jongenburger. 2001. 'Elicitation techniques in a Dutch syntactic dialect atlas project.' In: H. Broekhuis & T. van der Wouden (ed.), *Linguistics in The Netherlands* 2001, 18. John Benjamins, Amsterdam/ Philadelphia, 57-69.

Gerritsen, Marinel (ed.) (1991). *Atlas van de Nederlandse dialectsyntaxis* (AND). Amsterdam: P.J. Meertens-Instituut voor Dialectologie, Volkskunde en Naamkunde.

Gerritsen, Marinel (1993). The methodology of the Syntactic Atlas of Dutch. In: *Historical dialectology and linguistic change. Linguistic atlases and dictionaries*. Stuttgart: Franz Steiner Verlag. p 343-367.

Goeman, A. & J. Taeldeman (1996). Fonologie en morfologie van de Nederlandse dialecten. Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. In: *Taal en Tongval* 48-1, 38-59.

Goosens, J. (1977). *Inleiding tot de Nederlandse Dialectologie*. Wolters-Noordhoff, Groningen.

Greenbaum, Sidney (1973). Informant elicitation of data on syntactic variation. *Lingua* 31. 201-212.

Guy, Gregory R. (1980). Variation in the group and the individual: the case of final stop deletion. *Locating language in time and space*, ed. by David Sankoff and William Labov, 1-36.

Hinskens, F. (1996) *Dialect Levelling in Limburg: Structural and Sociolinguistic Aspects.* Linguistische Arbeiten, Tübingen: Niemeyer.

Jongeneel, Jacob (1884). *Dorpsspraak van Heerle vormenleer en woordenboek*. Heerlen: Van Hooren 1980.

Kroch, Anthony, S. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199-244.

Labov, William (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

¾¾ (1996). When intuitions fail, *Papers from the 32nd Regional Meeting of the Chicago Linguistics Society*, 32: 76-106.

Milroy, L. (1987). *Observing and analyzing natural language*. Oxford/New York: Basil Blackwell.

Oxford, R.L. (1982). Technical issues in designing and conducting research on language skill attrition. In: R. Lambert en B.F. Freed (red). *The loss of language skills*. Massachusetts: Rowley, 119-137.

Radford, Andrew (1997). *Syntactic theory and the structure of English*. Cambridge: CUP.

Winford, D. (1996). The problem of syntactic variation. In: J. Arnold, R. Blake, B. Davidson, S. Schwenter & J. Solomon (red.*). Sociolinguistic variation. Data, Theory, and Analysis. Selected Papers from NWAV 23.* Stanford California: CSLI Publications. 177-192.

▪ APPENDIX

Test sentences (cf. Cornips 1994: 82, 245-246):

A Infinitival purpose clauses, standard Dutch version[9]

(i)	en	die	visjes	zijn	niet	voor *om*		op	te	eten
	and	those	little fish	are	not	COMP		PART	to	eat
(ii)	Is	tien uur	voor	Klaas	nog	niet	te	laat	*om*	naar
	is	*ten o'clock*	*for*	*Klaas*	ADV	*yet*	*too*	*late*	COMP	PREP
	de	film	te	gaan?						
	the	*movie*	*to*	*go*						
(iii)	Vroeger	zaten	er	()rode	ringen	*om*	de	vliegtuigen		te
	formerly	*were*	*there*	*red*	*rings*	COMP	*the*	*planes*		*to*
	waarschuwen.									
	warn									
(iv)	Als	de	kleinkinderen		komen		dan	kan	ik	()iets
	if	*the*	*grandchildren*		*come*		*then*	*can*	*I*	*something*
	voor	hun	kopen	*om*	mee	te	spelen.			
	for	*them*	*buy*	COMP	*with*	*to*	*play*			
(v)	Komt	dat	voetbalelftal	(...)	*om*	te	voetballen?			
	come	*that*	*football-team*		COMP	*to*	*play-football*			
(vi)	en	dan	kun	je	'm	niet	meer	voeren	*om*	aan	te
	and	*then*	*can*	*you*	*him*	*no*	*longer*	*feed*	COMP		*to*
	sterken										
	strenght										

'and then you are no longer able to feed him in order to build up his strength'

[9] Note that om (4) is an infinitival relative clause.

B Infinitival non-purpose clauses, standard Dutch version

(i)	Het	is	me	erg	lastig	*om*	al	die	zinnen	na	te
	it	*is*	*me*	*very*	*difficult*	COMP	*all*	*those*	*sentences*	*too*	
	zeggen.										
	repeat										
(ii)	( )dat	dat	te	mooi	is	*om*	waar	te	zijn.		
	that	*that*	*too*	*beautiful*	*is*	COMP	*true*	*to*	*be*		
(iii)	Het	is	onmogelijk	voor	de	politie	in	Heerlen	*om*		
	it	*is*	*impossible*	*for*	*the*	*police*	*in*	*Heerlen*	COMP		
	overal	op	te	letten							
	everything		*too*	*watch*							
(iv)	Het	is	goed	*om*	te	weten (...)					
	it	*is*	*good*	COMP	*too*	*know*					
(v)	Het	is	me	te	moeilijk	*om*	(...)die	zesentwintig			
	it	*is*	*me*	*too*	*difficult*	COMP	*those*	*twenty-six*			
	zinnen	in te spreken									
	sentences	*to speak*									