

Elicitation techniques in a Dutch syntactic dialect atlas project*

Leonie Cornips and Willy Jongenburger
Meertens Instituut

1. Introduction

In the SAND-project dialectological, sociolinguistic and generative syntactic researchers in the Netherlands and in Dutch-speaking Belgium join forces in order to investigate and visualise the geographic distribution of syntactic variation in the Dutch dialects. The project is demanding with respect to methodology since up to now there is no standard procedure for syntactic elicitation techniques in dialect research with respect to large atlas projects. Hence, in several 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. This corresponds to atlas projects all over the world in which no more than five percent of published dialect maps involve syntactic data (Gerritsen 1991:9). Recently, several syntactic dialect atlas projects are conducted in Europe, for instance, the North Italian and Swiss German dialect atlases (cf. Bucheli & Glaser, to appear).

In this paper we will focus on different elicitation techniques that are used for the written and oral data collection undertaken by the SAND-project. First of all we will outline the different stages in the design of the project and their respective goals. Then we will discuss the results of several written and oral tests and finally we will draw conclusions about the workability and validity of the tests.¹

2. Phases of planning the data collection

Step 1: Literature study

First of all we started with a comprehensive literature study. All publications, i.e. articles and dialect grammars that have appeared on Dutch dialect syntax were

traced and all titles were fed into a database on the internet. At the moment this bibliography contains 1310 titles.² On the basis of the syntactic phenomena already described in the (dialect) literature a written questionnaire was prepared.

Step 2: Written questionnaire

The second step in the design of the SAND-project is an extensive written questionnaire containing 424 questions (including sub-questions and remarks to be made by the informants) that was sent out to 850 informants of the Meertens Instituut. In this questionnaire several tasks or elicitation techniques were used in order to be able to investigate the distribution of syntactic variables in a systematic way. These tests and their validity will be discussed in Section 3. The goal of the questionnaire is fourfold. First, the responses to the questionnaire provide insight into the geographic distribution of the syntactic variation investigated. Secondly, a systematic methodological analysis of the responses sheds light on the validity of each type of (written) elicitation (cf. Oxford 1982). Thirdly, the responses will show which part(s) of the Dutch-speaking area are of particular interest with respect to syntactic variation. Finally, the results of the written questionnaire will be extremely useful as input for the next phase: the oral fieldwork. Note that the data obtained with the written questionnaire will not be used directly for the atlas. The practical utility of the data obtained with written elicitation techniques is entirely restricted to the preparation of the oral fieldwork.

Step 3: Oral fieldwork

On the basis of the (geographic) distribution of the syntactic variables (cf. Cheshire 1987 for a definition of the notion syntactic variable) determined by means of the written questionnaire, interviews with dialect speakers are to be prepared for 250 locations in the Dutch-speaking area. For each syntactic variable to be investigated an appropriate elicitation test is prepared, resulting in a global protocol including all syntactic phenomena. Phenomena that do not occur in a specific area can be excluded from the interview protocol for that specific area, resulting in several geographically constrained protocols. In Section 4 the results of oral elicitation techniques used in seven pilot interviews with dialect speakers will be discussed.

Step 4: Visual presentation

The syntactic data gathered by means of the interviews with dialect speakers will be fed into a database. Linking this database to a sophisticated cartographic system will enable to visualise the complex of syntactic phenomena. A lot of syntactic phenom-

ena are not binary, for instance verbal cluster interruption may appear with particles, prepositional objects, negation, temporal adverbs, indirect objects and even subjects. Thus, the database and cartographic system must be highly flexible as the number of potential maps is unlimited. Note, that this cannot be done in the traditional atlases consisting of a fixed number of maps (cf. Gerritsen 1991; Goeman & Taeldeman 1996). Of course, the use of the cartographic database does not exclude the possibility of publishing several annotated maps. The syntactic cartographic system may prove to be an important research tool in generative syntax, sociolinguistics, dialectology and corpus linguistics.

3. Written elicitation

Table 1 indicates the response rates for the written questionnaire which initially sent out to 850 native speakers of diverse local dialects throughout The Netherlands and the Dutch speaking part of Belgium. A non-response of 52 informants is due to the fact that the addressees of the questionnaire were no longer able to complete the questionnaire due to certain mitigating factors such as age or passing away.

Table 1. The (non-)response of the written questionnaire

the questionnaires totalled	850
minus non-response (too old, sick, passed away, etc.)	798 (100%)
received questionnaires	368 (47%)

The 798 native speakers of the local dialects surveyed are 'professional' informants, that is to say, these speakers complete a written questionnaire from the Meertens Instituut every year. It is for this reason, in addition to the fact that the results of the written questionnaire are only of interest to the preparation of the oral fieldwork, that this questionnaire is so extensive (424 questions). In general, such a length promotes a very high non-response, i.e. the majority of the informants will not send back the questionnaire or they will fill it in only partially (see later). All questionnaires are coded and stored into a database (Filemaker).

Overview: Elicitation tasks and non-response

A written questionnaire 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 such as: (i) the repetition effect (cf. Bock 1986); (ii) the judging of syntactic constructions as ungrammatical simply on the basis of lexical items or 'knowledge of the world' (cf. Greenbaum 1973:210) and (iii) the fact that written forms are unduly influenced by prescriptive educational practices. It is for this reason that written data obtained will not be used directly for the atlas unlike the corpus obtained from oral fieldwork. Nevertheless, written elicitation data, provided these are interpreted carefully and considered in any resultant analysis, can be useful indicators of dialectal preferences at the syntactic level as demonstrated in Cornips and Corrigan (2001).

Throughout the questionnaire, the native speakers were asked *indirectly* about their grammaticality judgments. Thus, questions took the form 'do you ever encounter this variant in your local dialect?' rather than asking *direct* questions: 'do you consider this sentence (un)grammatical or (un)acceptable' or 'do you consider this sentence (in)correct in your dialect'. This is because direct questioning is thought to be one of the factors that promotes false responses — often in the direction of a socially superordinate standard norm which takes precedence over the native dialect system (Labov 1996:100). The answers to direct questioning methods may reflect the (syntactic) variant which the informants believe has prestige or is "correct" rather than the form they actually use (Labov 1972:213).

In general, the questionnaire is offered in Standard Dutch although some questions accommodate to a specific geographical distribution of the variant by inclusion of specific local dialect features such as subject-doubling, as illustrated in (1). In some dialects we encounter sentences such as:

- (1) Misschien ga 'k 'et (e) (k) ik wel krijgen
 Maybe go I it (I) got
 Betekenis: Misschien ga ik het wel krijgen
 Meaning: 'maybe, I'll get it'

- (i) Do you encounter sentences such as (1) in your local dialect?
 answer: yes or no

In this project, it was thought to be relevant to also administer constructions which, as far as we know, are ungrammatical in the Dutch dialects. The task exemplified in II below is a good example of this; it offers all possible orders including those that are deemed to be strictly ungrammatical in the verb cluster that contains three verbs.

The written questionnaire includes the tasks exemplified below viz. indirect grammaticality judgements (I.) & (II.); translation (III.) & (IV.); empty spots (V.). Below we will discuss these various elicitation tasks and their non-response. We define the notion non-response as those cases where the informant did not fill in *any* answer. The analysis of non-response may shed light on grammaticality aspects of the syntactic variants in question. It is important to note that we are not able to interpret the

written test data with material based on the subject's actual speech patterns due to the large geographical area and the density of the grid (250 locations).

- (I) *indirect grammaticality judgements* of the type:
 'do you encounter this variant in your dialect'; answer: yes or no

In the written questionnaire 61 items of this question type were included. The mean non-response of this question type is 20, meaning that from the 368 questionnaires, on average 20 were returned without an answer to items of this specific question type (5%). The non-response dispersions between 9 (2%) and 35 (10%). The following item evoked a non-response of 10%:

- Do you encounter this variant in your dialect? answer: yes or no
 Speaker A: Er komt morgen iemand langs.
 'Tomorrow someone will pay a visit'
 Speaker B: Wie dat?
 who that

Detailed examination of the answers reveals that the grammaticality of this specific construction is geographically concentrated in Belgium including the southern part of Dutch Limburg, and the northern part of the Netherlands, namely Groningen, Friesland and Drenthe. Analysis of the non-response shows that a relatively high number of non-responses occurs if the sentence is to a certain degree unacceptable in the local dialect. If it is grammatical, the informant consistently replies with 'yes' but if the construction is ungrammatical, the informants do not tick the answer 'no' and as such reveal a non-response. Apparently, the answer 'no' requires more (mental) activity for the informant than the answer 'yes'.

- (II) *indirect grammaticality judgements combined with a five points scale*; the informants have to indicate how uncommon (highest value = 1) or how common (highest value = 5) the order within the verbal cluster in their local dialect is:

- | | <i>encounter</i> | <i>uncommon-common</i> |
|--|------------------|------------------------|
| (2) | | |
| a. Ik weet dat Jan hard moet kunnen werken | yes/no | 1-2-3-4-5 |
| b. Ik weet dat Jan hard moet werken kunnen | yes/no | 1-2-3-4-5 |
| c. Ik weet dat Jan hard kunnen moet werken | yes/no | 1-2-3-4-5 |
| d. Ik weet dat Jan hard kunnen werken moet | yes/no | 1-2-3-4-5 |
| e. Ik weet dat Jan hard werken kunnen moet | yes/no | 1-2-3-4-5 |
| f. Ik weet dat Jan hard werken moet kunnen | yes/no | 1-2-3-4-5 |
| I know that Jan hard, work, must, can | | |

In the questionnaire 89 items of this question type were included. The mean non-response of this specific type of elicitation is 21 (6%) and ranges between 10 (3%)

and 41 (11%). The construction involving a verb cluster that contains an unacceptable word order of three verbs in 2e above induces the highest non-response of all.

(III) *translation*: will you translate the following sentence(s) in your dialect?

In the questionnaire 145 items of this elicitation type were included. The mean non-response of this specific task is 27 (7%) and ranges between 10 (3%) and 103 (28%). The translation of the ungrammatical Wh-question (partial) in (3) reveals the highest non-response of 28%.

(3) *Wie denk je die ik in de stad ontmoet heb?*

Betekenis: *Wie denk je dat ik in de stad ontmoet heb*

Meaning: *who think you that I in the town met have*

Vertaling (translation).....

Geographically, the non-response in this particular task is concentrated in the area South-Holland-Utrecht. By contrast, the Dutch provinces Gelderland and Overijssel and Belgium with the exception of the province of Limburg show hardly any non-responses. Apparently, the dialect speakers in these areas are able to interpret (3) quite readily.

(IV) *translation combined with ticking*:

The informant has to translate and to tick one out of two possible interpretations e.g. double negation or negative concord interpretation in (4a) and (4b):

(4) *Piet denkt dat Jan en Marie op niemand niet boos zijn*

Piet thinks that Jan and Marie at no one not angry are

Mogelijke betekenis (possible meaning):

(4) a. *Piet denkt dat Jan en Marie op iedereen boos zijn*

'Piet thinks that Jan and Marie are angry at everyone'

b. *Piet denkt dat Jan en Marie op niemand boos zijn*

'Piet thinks that Jan and Marie are angry at no one'

This type of elicitation was represented by 23 items in the questionnaire. The mean non-response is 54 (15%) and ranges between 21 (6%) and 108 (29%). This is a high non-response in comparison to the other elicitation tasks, mentioned earlier. Closer examination reveals that it is likely that this high rate is due to the meaning of the sentences in (4) that appeared to be very difficult to interpret (see also sentence 8 in Section 4).

The sentence in (5) with sentence final *niet* that closes off negative sentence evoked the highest non-response although the intended interpretation is simple clausal negation:

(5) *Er mag niemand spreken over dit probleem niet*

'No one is allowed to speak about this problem'

(V) *empty spots-task*:

the informant has to fill in the relevant function words from the local dialect e.g. in

(6) the infinitival complementizer and the position of *te*:

(6) *Heb je genoeg mensen... de mais van het land... halen?*

have you sufficient people ... the corn of the land... get-INF

'Do you have sufficient farm-workers to harvest the corn'

The non-response of this type of elicitation is irrelevant since empty spots may represent the grammaticality of the construction.

4. Oral elicitation

As mentioned above, the atlas will eventually be based on the data gathered during tape-recorded interviews. For the validity and consistency of the atlas it is very important that these are conducted in a consistent, reliable way. In Cornips & Jongenburger (2001) we argue from a methodological point of view that the interviews should be conducted in the local dialect rather than in the standard language or some regiolect, in order to avoid accommodation, i.e., adjustment from the dialect in the direction of the standard-like varieties. For most of the 250 locations selected this means that in order to be able to interview the informant in his or her own dialect, we are often occasioned to summon the assistance of another dialect speaker from the same community speaking the same variety (see Cornips & Jongenburger 2001 for a detailed overview of the design and the methodology of the SAND-project). To be able to evaluate the validity of the elicitation tests used, we conducted seven pilot interviews that will be discussed here.

Two of the pilot interviews were administered in de Randstad, the highly urbanised economic centre of the Netherlands; two others were conducted in Friesland, one was undertaken at the Betuwe in Gelderland and two were recorded in Vlaanderen (Vlaams Brabant). The results of all seven are discussed below by type of elicitation task.

Translations

In de Randstad, informants were reluctant to translate sentences in Standard Dutch into dialect, because the structural differences between the standard and their variety are very small. Or as one of our informants put it: 'Rotterdams is Standard

Dutch.' This problem could be solved by asking the informants how the sentence would sound in their dialect, drawing their attention from the lexical level to the level of pronunciation.

Indirect grammaticality judgments of the type: 'Do you encounter this sentence in your dialect'; answer: yes or no

See Section 3 for a detailed description of this test. This type of questions proved to be easy to perform and was, therefore, very useful in the interviews.

completion task

This task, where the assistant utters the beginning of a sentence and the informant has to finish it, was rather difficult. Instead of using the appropriate construction, some informants came up with totally irrelevant answers. An example of this type of elicitation is illustrated in (7) in which the phenomenon investigated is the position of the preposition *door* 'through':

- (7) Piet zegt: Zij liep de tuin door.
Piet says: She walks the garden through.
'Piet says: She walks through the garden.'

Marie antwoordt met: O, ik wist niet d _____

Marie answers with: O, I know not th _____

'Marie answers with: O, I did not know th _____'

Meaning questions

This task, where the informant is asked to give the meaning of a sentence that was uttered by the assistant, appeared to be far too difficult to perform. When the informants heard the sentence illustrated by (8), for instance, they could not make a motivated choice between the options — double negation or negative concord interpretation —:

- (8) Wim denkt dat we *nooit niemand* een prijs geven.

Wim thinks that we never nobody a prize give.

possible meanings:

- (8) a. Wim denkt dat we altijd iemand een prijs geven.

Wim thinks that we always somebody a prize give.

'Wim thinks that we always give somebody a prize.'

- b. Wim denkt dat we niemand ooit een prijs geven.
Wim thinks that we nobody ever a prize give.
'Wim thinks that we nobody ever give a prize.'

Repetition task

Long sentences uttered by the assistant were to be repeated by the informant. The sentences were extremely long so that the informant could not confine himself to repeating them literally, but had to fall back on his own (dialect) grammar to reconstruct the sentence (cf. Cornips 1994). This task appeared to be very useful for testing word order in verbal clusters since the informants adjusted word order in the repeated sentence according to their dialect grammar.

Finally, we want to make some general remarks with respect to the use of oral elicitation tasks. Instructions should be as short as possible, since longer instructions risk confusion on the part of the informant. Speakers seem to prefer a few longer sessions of one question with breaks rather than many short incessant sessions of different types. In other words, they prefer eight questions of one type in one large cluster to four questions in two smaller clusters.

5. Conclusions

In this paper we have discussed the results of several written and oral elicitation techniques that have been applied to the data collection programme associated with the SAND-project. It appears from this preliminary analysis that not all types of

Table 2. Type of elicitation and their usability in respectively a written and an oral setting

type of elicitation	Setting	
	written	oral
repetition task	*	+
indirect grammaticality judgments 'do you encounter?'	+	+
indirect grammaticality judgments on 5-points-scale	+	*
translation task		geographic area? +
empty spots test	+	*
completion task	**	-
meaning questions	+/-	-

Legend: + = task is easy to perform; - = task is difficult to perform; * = inappropriate setting for the type of elicitation involved; ** = not used in the SAND-questionnaire.

elicitation are workable in both written and oral settings. In Table 2 we recapitulate our experiences in this regard: '+' means that the specific task is easy to perform; '-' means that the task is difficult to perform. Some types of elicitations are typically written tasks, as is the case with indirect grammaticality judgements and the empty spots test. Other tasks, such as the repetition task, are only workable in an oral setting. In Table 2 these setting-specific elicitation techniques are marked with an asterisk (*) for the inappropriate setting. The two asterisks (**) appended to the completion task in the written setting indicate that this type of elicitation was not used in the written questionnaire (i.e. this task is not by definition unworkable in a written setting, it has simply not been used in this investigation).

Furthermore, we have shown that for some types of elicitation, the practicability of a task depends on the region investigated. In regions where the dialect varieties do not differ much from the standard language, i.e. in de Randstad, it is not useful to ask for written translations, whereas it is possible during interviews to ask how a sentence would *sound* in the same dialect.

Notes

* We would like to thank the fieldworkers/PhD-students of the project Jeroen van Craenenbroeck, Gunther De Vogelaer, Tamar Israel, Susanne van der Kleij, Marjo van Koppen, Annemie Neuckermans, Henk Wolf en Hedde Zeijlstra.

1. The notion validity refers to obtaining evidence that is clearly interpretable as being evidence of the behaviour we are interested in (cf. Spolsky 1988).
2. <http://www.meertens.knaw.nl/projecten/sand/sand.html>

References

- Bock, J.K. (1986) 'Syntactic persistence in language production'. *Cognitive Psychology* 18, 335-387.
- Bucheli, C. & Glaser, E. (to appear) 'The syntactic atlas of Swiss German dialects: empirical and methodological problems'. In S. Barbiers, L. Cornips and S. van der Kleij, eds. *Syntactic Microvariation*. Electronic Volume Meertens Instituut
- Cheshire, J. (1987) 'Syntactic variation, the linguistic variable, and sociolinguistic theory'. *Linguistics* 25, 257-282.
- Cornips, L. (1994) *Syntactische variatie in het Algemeen Nederlands van Heerlen* (Syntactic variation in Heerlen Dutch). Amsterdam: Dissertation, IFOTT 6.
- Cornips, L. & Corrigan, K. (2001, to appear) 'Convergence and divergence in grammar'. In P. Auer, F. Hinskens and P. Kerswill, eds. *The convergence and divergence of dialects in contemporary societies*. Cambridge University Press.
- Cornips, L. & Jongenburger, W. (2001) 'Het design en de methodologie van het SAND-project'. *Nederlandse Taalkunde* 2001-3, 215-232
- Gerritsen, M. (ed.) (1991) *Atlas van de Nederlandse dialectsyntaxis* (AND). Amsterdam: P.J. Meertens-Instituut voor Dialectologie, Volkskunde en Naamkunde.
- Goeman, A. & Taeldeeman, J. (1996) 'Fonologie en morfologie van de Nederlandse dialecten. Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten'. *Taal en Tongval* 48-1, 38-59.
- Greenbaum, S. (1973) 'Informant elicitation of data on syntactic variation'. *Lingua* 31, 201-212.
- Labov, W. (1972) *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1996) 'When intuitions fail'. *Papers from the 32nd Regional Meeting of the Chicago Linguistics Society*, 32, 76-106.
- Oxford, R.L. (1982) 'Technical issues in designing and conducting research on language skill attrition'. In R. Lambert and B.F. Freed, eds. *The loss of language skills*. Massachusetts: Rowley, 119-137.
- Spolsky, B. (1988) 'Tests in sociolinguistics'. In: U. Ammon, N. Dittmar and K. Mattheier, eds. *Sociolinguistics*. Berlin, New York: De Gruyter.