ELSEVIER

On standardising syntactic elicitation techniques (part 1)[☆]

Leonie Cornips[a,*], Cecilia Poletto[b]

[a]*Meertens Instituut (KNAW) Amsterdam, The Netherlands*
[b]*ISTC, CNR, Padua, Italy*

Abstract

The aim of this article is to provide a first overview on several syntactic elicitation techniques used by dialectologists and sociolinguists and to evaluate them from a generative perspective. Both oral and written techniques will be taken into account and analysed according to the sociolinguistic situation. In addition, basic problems like the choice of subject, the language used for the interaction will be addressed and discussed on the basis of two syntactic Atlases, the Dutch *SAND* and the Italian *ASIS*.
© 2004 Elsevier B.V. All rights reserved.

1. Introduction

The role of syntactic variation in closely related dialects, or the domain of microvariation, is receiving a growing interest in theoretically oriented research. Microvariation studies are primarily concerned with: (i) the possibility of testing potential correlations between syntactic variables and (ii) analysing the distribution of a single variable while

---

[☆] Part 2 will deal with the relevance of methodology in theoretical research and will discuss a case study on negation and particles.

[*] Corresponding author. Present address: Meertens Institute, Toan Muyskenweg 25, P.O. Box 94264, 1090 66 Amsterdam, The Netherlands.

E-mail address: leonie.cornips@meertens.knaw.nl (L. Cornips).

keeping other, potentially interfering factors constant.[1] The study of a single variable requires that all other extraneous factors are recognised and controlled for. It is clear that, if we intend to provide a solid empirical basis for studying the minimal directions in which dialects can vary, the role of field work and the procedures through which data are elicited and ordered becomes not only central but also extremely delicate.

This paper has two main goals: (i) to propose a standardisation of the design and implementation of grammaticality tasks and, (ii) establishing the manner in which non-linguists should be interrogated about the (un)acceptability of sentences in their dialects. Hence, this paper will draw on a range of previous research in order to explore issues related to the use of various elicitation techniques which are thought to be valid research tools.

Syntactic dialect[2] atlas projects, in particular, have to deal with valid methods for obtaining reliable and sufficient syntactic data from dialects across a large geographical area. Examples of syntactic dialect atlas projects that are currently undertaken in Europe are the Northern Italian syntactic dialect atlas (ASIS) and the syntactic atlas of the Dutch dialects (SAND) (see Barbiers and Cornips, 2002).[3] These syntactic dialect atlas projects are very unique with respect to the object of research. Hence, in traditional dialect maps globally, no more than 5% of those that are published involve syntactic data and this is purely chance. Moreover, the data are not organised in a way that permits immediate comparison and analysis of the syntactic information contained in the maps. Similarly, traditional dialect monographs usually do not provide information about the syntax, which can only be indirectly recovered in some cases through the description of morphological patterns.

It is not just the geographic reach that makes ASIS/SAND different from other types of syntactic research. Another major characteristic of them that is unusual is their aim, i.e. not to provide a simple description of the syntactic phenomena investigated. Rather the data collection task in each case is seen as instrumental to a deeper understanding of the I-language of the speakers. Although we usually conceive language as a single object of scientific study, microvariation clearly shows that each speaker has his/her own I-language, which minimally differs from the I-language of everybody else, including people educated in the same environment at the same time. We could conceive the study of language therefore as similar to the study of families of bacteria in biology: the object of our investigation are families of very closely related languages, which, at first sight, appear to be identical, though they are not when subtler tools of inquiry are used. If I-languages are

---

[1] One clear example of correlation between two syntactic variables comes from the the distribution of subject clitics in the Northern Italian dialects. The form and the syntactic properties (possibility of enclisis in main interrogative structure, cooccurrence with a quantified subject, etc.) of subject clitics are different when the inflected verb, that they are adjacent to is a main or an auxiliary verb. If this connection between type of subject clitic and type of verb is not recognized and the data concerning both types are put together, the resultant findings are uninterpretable and apparently non-coherent.

[2] The term 'dialect' can be treated as a purely relational concept, which means that, by definition, without a standard there can be no dialect. In addition, the term dialect is also used exclusively in order to refer to areal variability within a language (cf. Auer). Of course, dialects do not derive in any sense from the standard languages, but standard languages are simply dialects that have gained a different social prestige and have undergone a special process of lexical enrichment or codification and are used in higher stylistic levels and writing.

[3] More information about the ASIS- and the SAND-projects can be found at: http://asis-cnr.unipd.it and http://www.meertens.nl/projecten/sand/sandeng.html, respectively.

all different, E-languages, although very similar, are also minimally different, once we look at them closely enough. As in biology it is necessary to take a microscopic approach and, as such, studying microvariation requires the development of new research tools. It is not surprising then that a by-product of the ASIS/SAND enterprises has been the development of new methods for the elicitation of syntactic data. Usually, studying the I-language of speakers involves creating various experiments in the form of acceptability judgement tasks. Because of the type of investigation involved, these have to be systematically elicited from a sample of community members in a large geographical area rather than being derived from linguists' own introspections. Although the focus is on native-speaker introspection, it is certainly not the case that this is done in an idealised environment. Hence, in this empirical domain, the opposition between I-language and E-language phenomena is not, necessarily, watertight (Milroy, 2001; Muysken, 2000).

This paper is organised as follows. In the first part, we will focus on two general issues: (i) the relation between intuitions and acceptability judgements and (ii) the mismatch between acceptability judgements and the actual use of constructions by speakers. In the second section, we will discuss various grammaticality tasks with their possible task-effects. In addition, we will propose a solution to the design of grammaticality judgement tasks such that they can be designed so as to enhance their validity and reliability. The last section focus on the elicitation of what one might describe as complex constructions.

2. The relation between intuitions and acceptability judgements

Although much recent work in dialectology[4] almost exclusively make use of data gathered by recording spontaneous speech, there are several reasons for using other techniques, especially eliciting grammaticality judgements (Schütze, 1996: 2). 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. In addition, by eliciting acceptability judgements we can examine reactions to sentence types that might occur only very rarely in spontaneous speech or recorded corpora.

Further, we are able to elicit syntactic variables that do not always show up in interaction with other relevant syntactic variables in spontaneous speech, but that are predicted by theory to do so. For instance, it has been noted that in some Northern Italian dialects (Trentino), the doubly filled comp filter can be violated (yielding the sequence *wh*-complementizer-verb) only when the subject is found in postverbal position. Checking this hypothesis is only possible by eliciting ungrammatical judgements for sentences with the sequence *wh*-complementizer and subject.

Finally, we are able to focus on possible negative data and also derive a scale of grammaticality concerning a given phenomenon, which would never arise by the simple observation of corpora. Observational studies, as is the regular practice in variationist investigations, do not always provide a high enough concentration of the phenomena we are interested in. In most cases the phenomenon under investigation is intertwined with other variables which might interfere in its definition; in other words, we cannot control the

---

[4] See among others, Sanga (1991).

variables that are interacting in our sample. As a result, in addition to the 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. Hence, a truly experimental approach towards linguistic phenomena is required, where special tasks are designed to test the distribution of a single phenomenon or the interaction between two phenomena. The observation of spontaneous speech does not correspond in any sense to the type of controlled environment needed for an experiment of the sort used in, say, physics or chemistry.

However, there is generally agreement today that grammaticality judgement tasks do not, necessarily, provide a direct window into an individual's competence alone; other factors may influence the character of the data elicited (Altenberg and Vago, 2002).

According to Schütze there are, in fact, good reasons to assume that judgements also involve performance. After all, in order to assess a sentence, one presumably has to first process it, or attempt to process it; thus, at least some of the performance factors involved in normal sentence processing ought to play a role in grammaticality judgment tasks. Chomsky (1986: 36) argues that: *"In practice, we tend to operate on the assumption, or pretense, that these informant judgments give us "direct evidence" as to the structure of the I-language, but, of course, this is only a tentative and inexact working hypothesis. In general, informant judgments do not reflect the structure of the language directly; judgments of acceptability, for example, may fail to provide direct evidence as to grammatical status because of the intrusions of numerous other factors"*.

Furthermore, it is generally accepted that acceptability occurs on a continuum presented as **, *, ?*, ??, ?. As research has proceeded, data have become increasingly subtle and empirical adequacy more and more difficult to obtain (Gervain, 2002). According to Gervain (2002), it is important to note that degrees of acceptability are not in themselves problematic. What is controversial here is their imprecise treatment[5]. It is not clarified what scale, if any, is being used, how the different degrees relate to each other and how they get interpreted in the analysis (e.g. is a two question mark sentence a piece of evidence or a piece of counter-evidence?). Moreover, each subject might interpret the scale rather differently when testing the data.

In conclusion, on the one hand we cannot use spontaneous speech alone to study the distribution of linguistic phenomena; on the other, eliciting grammaticality judgments might pose some problems, especially (as we will see) in a bilingual[6] environment in which the use of dialect is also constrained by sociolinguistic factors.

2.1. The mismatch between the judgments about a construction and its use

The fact that a native speaker judges a certain form to be completely unacceptable, but can, nevertheless, be recorded using it freely in every-day conversation, is a striking result

---

[5] Moreover the comparison set is important and extremely difficult to establish.

[6] We refer to the notion bilingualism or bilingual speaker in its broadest sense, i.e. to include contact between speakers of different languages as well as contact between speakers of different dialects as speakers between dialects and standard languages.

of elicited introspective judgments (Labov, 1996: 78). A distinction that is essential here is that between explicit and implicit knowledge. Paradis (1994), linking implicit knowledge and competence, views implicit competence as knowledge that individuals themselves are not aware of (taken from Altenberg and Vago, 2002). It is further proposed that implicit knowledge is not conscious and is used automatically, and that explicit knowledge is conscious knowledge that has been learned primarily in school. It is assumed that acceptability judgment tasks cannot rely entirely on explicit knowledge since native speakers are able to make judgments about structures with no explicit knowledge about them and which had not explicitly been taught to them. On the other hand, it might be the case that in producing acceptability judgments people tend to observe prescriptive grammar rather than rely on actual usage. Also, acceptability judgments may be based on estimated frequency of usage, or on the degree of semantic or pragmatic plausibility.

Strong evidence for the preference of the prescriptive norm is presented by socio-linguistic research. Since Labov (1972), the unreliability of native-speaker judgments is well known. He has shown that: ''whenever a subordinate dialect (stigmatised) 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. The problem of intuitive responses to these syntactic questions is acute: we do not know why speakers find it so difficult to recognise their native grammatical patterns. 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 varieties of the same 'language' differ with respect to social prestige as do the English dialects observed in the United States, and in those European cases in which the standard variety has emerged out of local dialects for historical reasons. For example, the influence from the standard variety is to be expected in the situation of dialect loss where the dialect is increasingly infiltrated by standard features. According to Auer, the breaking away of the most local, most dialectal forms is a matter of degree, and it can gradually affect large parts of the continuum between dialect and standard variety until almost nothing is left but the standard variety, as is the case in the Randstad area[7] of the Netherlands (cf. Auer). In this case, the standard variety will often be the most prestigious since it is aspired to by speakers of more than one dialect (which does not necessarily imply that it is mastered by everybody), and it is subject to attitudinal factors (cf. Milroy, 2002). What is more, it is not the fact of codification which makes a standard variety but the fact that its speakers think that the existence of a grammar and a dictionary should exist and that, where they exist, they should determine how members of that society ought to express themselves in situations in which the standard is required (cf. Auer).

---

[7] The Randstad area involves the cities Amsterdam, Rotterdam, The Hague and Utrecht. It is located in the western part of the Netherlands and it is the economical, administrative and cultural centre. The Dutch standard language originates from the local dialect of Amsterdam in the 17th century.

It is argued that the degree of interference of the prescriptive norm varies depending on the attitude of the speakers towards the two (or more) dialects they master. When the speakers are conscious that they are bilingual, and that their dialect is not the standard variety (but a different system), they behave in different ways depending on whether a given phenomenon exists only in one of the two languages or in both and whether a phenomenon is obligatory or optional. There seems to be two types of behaviours when a given phenomenon is tested: those phenomena that are obligatory in a given dialect are generally provided by native speakers even if the same phenomenon is banned from the standard variety. For instance, many Northern Italian dialects require a complementizer after the *wh*-item in embedded (or also in main) interrogatives[8]; this obligatoriness is never subject to interference from the standard language: the speakers always use the complementizer even though it is not allowed (indeed, prohibited) in the standard. In this case native speakers seem to be able to distinguish whether a given construction is grammatical without interference from prescriptive norms. However, this is not so for *wh* in situ in the varieties in which the phenomenon is optional. In many cases the speakers tend to reproduce the standard model, because this is nonetheless grammatical in their dialect. In other words, it seems that, at least in the Northern Italian dialects, the speakers permit interference from the standard only when grammatical principles of the dialect are not violated. This is especially true for the cases of 'optional' constructions, where interference from the standard language is active and prescriptive norms tend to induce in the speakers a preference for the standard construction, as we will discuss later. We see here that, even if grammaticality judgments do not resort to explicit knowledge, being aware of the existence of mastering different linguistic systems is crucial in determining the attitude of the speakers towards the judgment tasks.

Thus, adult responses on acceptability judgment tasks rely at least in part also on explicit, prescriptive notions held by speakers. So, the fact that the use of a construction does not necessarily imply its acceptance can be attributed to explicit, prescriptive knowledge about the superordinate variety. One way to diminish this effect is to ask for indirect acceptability judgments. Moreover, the seeking of (relative) judgments (see also Section 3.1) in an indirect way seems very felicitous according to the successful experimental methods described in Labov (1975). Hence, rather than eliciting direct intuitions by the formula: 'do you judge X a grammatical/better sentence than Y?', speakers can be asked the more indirect: 'do you ever encounter the variants X and Y in the local dialect?' and if so 'which variant Y or X do you consider to be the most or the least common one in your local dialect?' or 'which variant X or Y 'sounds' better in your dialect'. This first type of question 'do you ever encounter . . .' proved to be easy to perform and they are very useful for the subjects to accustom themselves to the test situation. However, the answer 'yes' or 'no' should not be considered as hard evidence for the (un)acceptability of the construction. Without further information, it is not certain that the

---

[8] For a detailed discussion of this feature see Poletto (2000). Here we report only an example:

(i)	a	*Cosa*	*che*	ti	vol?	Venetian	
		what	that	you	want?		'What do you want?'
	b	No	so	*cossa*	*che*	ti	vol
		Not	know	what	that	you	want 'I do not know what you want'

reasons for which subjects reject sentences are relevant to theoretical, syntactic issues at hand. Such elicited 'intuitions' differ from absolute acceptability judgments in that heterogeneity is assumed by providing several alternatives, that is to say, they are designed to reveal competence in bidialectal, heterogeneous varieties rather than in a single, homogeneous variety (cf. Cornips and Corrigan, to appear, Rickford, 1987).

2.2. Speech repertoires and acceptability judgments

From the above, it is obvious that, especially in a bilingual situation, information must be gathered about the specific speech repertoire of the dialect and standard variety to exclude mechanisms of social intervention or explicit, prescriptive knowledge as much as possible. In doing so, we can predict, more accurately, when explicit knowledge or prescriptive norms, such as socially superordinate forms, influence the native system and second, whether there is a relation between the social context and the (un)acceptability of grammatical judgments. Knowledge about speech repertoires in a social context minimises the risk that we obtain information about prescriptive norms of the standard or prestigious variety while our intention is to question the speakers about their dialect variety.

It is clear that different areas in Europe may behave differently with respect to the language of instruction in the written questionnaires and interaction in the oral elicitation tasks according to their speech repertoires. Following Auer, The Netherlands (excluding the Randstad area and Flemish-speaking Belgium) reveal a diaglossic linguistic repertoire which is characterised by intermediate variants between standard and (base) dialect. What we find in these areas is that, most usually, these intermediate forms may occur since there is no clear-cut separation of standard and dialect. Speakers can change their way of speaking without a clear and abrupt point of transition between dialect and standard. In the language of instruction and interaction in oral interviews in the Netherlands and in Belgium it is of major importance therefore to avoid standard variants and/or intermediate variants as much as possible; however, this is not the case for the Northern Italian dialects. The speech repertoire of speakers of the northern Italian dialects and the standard variety is not a bilingual one but contains a structurally related standard variety in addition to the dialects. Pellegrini (1982) arrives at distinguishing seven distinct 'stylistic levels' from the base dialect to the standard variety, all structurally related and used in different contexts. For instance, truncated past participles, which are typical of the dialect, can also be used in the spoken language at low stylistic levels, but are not admitted in high stylistic levels. The same is true of phrasal verbs, which do not exist at all in the standard Italian spoken in the central and southern regions.[9] Still, phrasal verbs are tolerated to a higher extent, they appear at all levels except the highest one, where the language is very formal.

Moreover, in the Northern Italian domain there is a considerable difference between those regions in which a regional dialectal variety has developed (for historical reasons of unity) and those where this has not happened. In the regions where a prestige dialect is present (e.g. Veneto, Piedmont, etc.), this generally strongly interferes with elicitation of

---

[9] Cases of phrasal verbs are particularly widespread, Veneto speakers use forms like *dir su* 'tell off', *mangiare fuori* 'eat up', *buttare fuori* 'through up', when they speak Italian. These forms are completely incomprehensible to non-Veneto speakers.

data in the local variety. Therefore, the preferred language for oral interviews is para-doxically the standard variety, which is felt "too far" from the local variety to interfere. If the regional variety is used, interference is much more easily triggered. Using the local dialect is always the best option, but for practical reasons it is obviously not always possible. In other words, influence from the standard variety is reduced in the northern Italian domain because there is another system, the prestigious non local dialect, related to the local dialects. So, it appears that, with respect to the northern Italian dialects at least, influence of the standard language is apparently very restricted.

3. The design of grammaticality tasks

3.1. The exclusion of possible interfering factors causing variation

In the design of dialect atlas projects, all other factors which can potentially influence syntactic variation have to be kept invariable as far as practically possible. In the Labovian framework, it has been shown in numerous urban studies that social dimensions of speakers correlate closely with patterns of language variation. Homogenising the sample with respect to the social profile or variables of the subjects will minimise to a great extent the risk of finding syntactic variation that cannot be attributed to geographical factors alone but also to social factors. Relevant social dimensions of speakers that are considered to be potential determinants of language variation are, social class, gender, age and ethnicity.

In the SAND-project, the following social variables were homogenised, as much as possible: (i) all subjects are native speakers of the local dialect; (ii) both the subject and their parents were born in the same community and have lived there until adulthood; (iii) the subjects did not leave their community for longer than seven years; (iv) the subjects speak their local dialect in several functional domains; (v) the subjects belong to middle–low level employees and (vi) the subjects are aged between 55 and 70 years.[10]

The ASIS project also controls for the above variables except for age since young speakers have also been tested, to reveal whether there are diachronic changes occurring in the phenomena under investigation. Moreover, in a subset of subjects data concerning grandparents have additionally been gathered to test whether their origin also plays a role or not.

Who is a good subject? Good subjects are those who are able to focus on the syntactic level and on their dialect, avoiding possible interference from the standard on the one hand or from some idealised form of more conservative dialect on the other. In the SAND-project, subjects are selected and questioned by telephone about their views on their dialect. Subjects who consider, for instance, the dialect of younger people or the dialect as

---

[10] Clearly from the point of view of generative grammarians it is not important which grammar is chosen, as any grammar is potentially interesting (in the ASIS project there are subjects of every age group, but the age of the subject is always known). However, due to the general process of standardization of the dialects, the age range selected for the SAND project is the one that has the highest probability of showing a dialect which is maximally different from the standard language. An anonymous reviewer points out that there might be no need to control for age in some cases, as this is the only age-group where there is dialectal competence. This is not the case for the two projects illustrated here.

it is today, to be 'wrong' are excluded. Furthermore, subjects are evaluated and those subjects who always copy the standard construction into their local dialect are excluded (cf. Carden, 1976).[11]

In the ASIS-project good subjects are able to give the interviewer all the possible alternatives for the constructions that are presented to them and, if it is the case, to connect the different alternatives to different contexts (which can reveal semantic distinctions). In order to do so, a subset of the subjects is also trained to become collaborators, who present alternative structures spontaneously. In such cases subjects also spontaneously provide information on their own and neighbouring dialects, which has proven valid. Further, the subjects are evaluated as a 'valid' or non-valid' on the basis of (i) the number of misinterpretations found in the text, (ii) the number of sentences they left out, (iii) the accuracy in proposing variants for the same sentence and providing possible semantic or pragmatic differences connected to the syntactic variants and (iv) coherence in the grammatical system.[12] Finally, with respect to the social variables of subjects, an interesting finding in the Swiss atlas-project is that the answers of a proportion of the highly educated people (mostly male), teachers and subjects interested in dialectology are unreliable in that they show normative judgments, hypercorrection and influences of the standard language (cf. Bucheli and Glaser, 2002). Often, subjects of this type are local poets who invent their own language in order to differentiate it from that which has been influenced by the standard: they reject phenomena that are usually present in the standard as well as in the dialect, interpreting them as an interference from the standard even when this is not the case. In more general terms, subjects of this type provide a falsified picture of the dialect.

3.2. A proper empirical method for assessing the acceptability of a construction

Every elicitation situation is artificial and the subject is being asked for a sort of behaviour that, at least on the face of it, is entirely different from everyday conversation. What are people, especially non-linguists, doing when judging the acceptability of a sentence? The only thing we are sure about is that we do not know exactly what they are doing (Schütze, 1996).

One of the methods (which seems to work) for obtaining acceptability judgments is to ask for relative judgments (see also Section 2.1). According to Schütze the speakers usually feel more confident about relative judgments than absolute ones. It is possible that speakers

---

[11] When a subject does not conform to the parameters listed in the following, his data are either excluded entirely or are checked with some other more reliable speaker of the same dialect. There can be various reasons why a speaker is not a good informant: some of them (for instance, local poets) have idealized their own dialect and trying to preserve it, they have sort of "reinvented" it, creating their own language. Other speakers are simply not very good at figuring out contexts and situations, or provide their own grammatical explanations for the data and then in the subsequent tests they try to remain faithful to this explanation, which can lead to totally misleading data.

[12] Every inquiry contains a number of apparently redundant data, which test the same variable under the same conditions: for instance, in the first ASIS questionnaire there are five sentences testing the possibility of having a postverbal subject with a transitive verb with a postverbal definite object. In general, informants provide the same type of (grammatical or ungrammatical) judgment for those.

always give relative judgments when asked about absolute judgments since they will compare constructions when considering their answer. Moreover, these indirect acceptability judgement tasks can easily be combined with a scale; e.g. the subjects have to indicate how uncommon or how common the variant is in their local dialect. [13] An example from a written questionnaire used in the SAND-project with respect to the order in the verbal cluster (right periphery) is presented in the following (Cornips and Jongenburger, 2001).

Instruction: Which variant do you consider to be the most common one in your local dialect. Please indicate how uncommon (highest value = 1) or how common (highest value = 5) the order is in your local dialect:

		Encounter	Uncommon–common
(a)	Ik weet dat Jan hard *moet kunnen werken*	yesno	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		

It is important to note, however, that relative judgments, as illustrated above, are not without problems in oral inquiries. One problem is efficiency that is, in particular, the amount of information one can extract from a given number of relative judgments is much less than the amount one can extract from absolute ratings. Therefore, in oral interviews the amount of alternative structures may be limited to two. Moreover, it might be the case that the distribution of alternatives is connected to semantic properties, and the speaker might find it difficult to say whether alternative (a) is more common than alternative (b).

The approach of requiring the subjects to rank order sentences from most to least acceptable has certain advantages (Schütze, 1996): (i) psychometric research[14] indicates that people are much more reliable on comparative, as opposed to independent ratings, (ii) rank orders also solve the problem of different baselines on a rating scale and there are non-parametric statistical tests for assessing the constituency or correlation between sets of rank orders. It appears to be an experiment with good success and it can be quantified very easily by means of a cluster analysis (Cowart, 1997).

However, it is important to note that a subject who judges alternative (a) better than (b) and also judges (b) better than (c) when considering two at a time, does not necessarily judge (a) better than (c) when they are examined side by side (cf. Schütze, 1996). Further,

---

[13] The five- or seven-point scale has to be changed if it is similar to the way of ranking in school systems in order to exclude any association with a formal test situation.

[14] Cf. on this topic Schütze (1996), Bard et al. (1996), and Cowart (1997).

for a naive judge, it has the defect of suggesting that both alternative (a) and alternative (b) and (c) are possible forms of the language.[15]

3.3. Oral and written elicitation

One positive side-effect of written elicitation is that the subject provides a first analysis of the constituting elements of the language in writing and separating them. One interesting example comes from Veneto dialects in which a subject clitic (*a*) appears in front of the verb *rivar* 'to arrive', which is 'arrivare' in standard Italian, so the two forms in Veneto and the standard are the following:

(1)   a   A riva
         b   0 arriva

In the dialect there is a subject clitic which cannot be found if the subjects do not write the sentence showing that the vowel *a* does not belong to the verb, but is an independent element. The same has been found with Southern Italian dialects that have object clitics and prepositions that form a unique cluster, which can only be analysed from the written form and not from oral data.

On the other hand, oral differs from written elicitation in that the former enables the researcher (i) to elicit a more natural reflection of ordinary language use and (ii) to observe and immediately respond to the reactions and answers of the subjects and (iii) to control for intonation and prosody with respect to topic and focus of the sentence. However, different levels of speech styles (informal and formal) may yield a complicating factor for both oral and written syntactic elicitation.

Spoken standard varieties may differ from written standard varieties, sometimes even considerably. There is at least one major reason why this might be the case. According to Auer, a spoken standard is subject, as all spoken varieties are, to the requirements of face-to-face interaction and on-line language processing under conditions of almost perfect speaker/hearer synchronisation. Structures which may be possible in spoken varieties do not always show up in writing/reading and vice versa. It is also the case that a written standard, once established, quickly becomes more conservative than the spoken standard. In the following, we will present two examples of how the written standard may differ from dialect or spoken regional varieties, as illustrated by the Bavarian (Weiß, 2002) and Heerlen Dutch examples in (2a) and (2b), respectively. Both spoken varieties show the phenomenon of negative concord, i.e. multiple occurrences of items overtly marked for negativity which do not cancel each other but form a single negation:

---

[15] One anonymous reviewer notes that this problem might be overcome by telling the subject that some sentences might be ungrammatical. However, some sentences might be more ungrammatical than others. One more serious problem might concern the fact that often the speaker is induced to judge as less good a sentence which is in fact less frequent in the sense that it requires a special semantic or pragmatic context to be elicited. This method thus imposes a considerable burden on the subject, namely, that of imagining possible contexts in which a give sentence might be uttered.

(2) a Bavarian		Gesdan	ha'e	*neamd*	*ned*	gseng.
					Yesterday	have-I	nobody	not	seen
	 b Heerlen Dutch Ik	heb	*nooit*	*geen*	Jägermeister meer aangeraakt.
					I		have	never	no		Jägermeister more touched

Typological research by Haspelmath (mentioned in Weiß, 2002) has shown that the pattern where only one item expresses negation (as in (3)) does not occur very frequently in natural languages: it is mainly restricted to some standard languages:

(3)	a	Nobody	came.
	b	Niemand	kam

According to Haspelmath and Weiß, there is good evidence that its development had to do with language external factors such as modelling languages after Latin grammar or logical considerations in the course of standardisation. So, the pattern in (3) can be considered to be a more artificial phenomenon. It is striking that all the dialects of the standard Germanic varieties such as English, German and Dutch reveal negative concord, whereas their written standards do not (of course, this is not the case for Romance languages).

Another example concerns the presence of complementizer agreement *dan* 'that', such as in the dialects of West-Flemish (Haegeman, 1992 taken from Vogelaer et al., 2002):

(4)	*Lapscheure*	Kpeinzen	da**n**-k	(ik)	morgen	goan.
					I-think		that-I	(I)		tomorrow	go

However, complementizer agreement will never be attested in the written standard variety. Bearing this in mind, although the method of written questionnaires has the advantage of systematically gathering dialect data in a large geographical area within a short time span, and although it is an elicitation technique that enables the researcher to standardise both the collection and the analysis of the material, this method induces the risk that people's responses are governed by prescriptive grammar and what we have described elsewhere as 'explicit knowledge' in the acceptability judgments that they produce. This is because written varieties are learned at school and most dialect varieties are not written but spoken.

Moreover, writing a language that is only oral might cause problems to the informants, who might be uncertain concerning the grapheme that he/she should use for phonemes that are only found in the dialect and not in the standard variety. It can happen that an informant refuses to write in the dialect just because of this. This might also happen for dialects which have a written tradition or, indeed, some standardised form used for local newspapers. The informant feels that he/she does not master that written system. However, once he/she is assured that the inquiry has syntactic purposes, the problem is, in general, overcome.

In conclusion, from a syntactic perspective, written and oral varieties may differ to a great extent. Constructions in spoken varieties do not always occur in written languages and vice versa. In addition, the written response of the speaker will be unduly influenced by prescriptive educational practices. Therefore, judgments on written language may be considered to be less accurate than judgments on spoken language when examining dialect varieties.

3.4. Language of instruction and of interaction

We already mentioned that the language of instruction and interaction is of extreme importance in eliciting the most base variant, i.e. dialectal variants. In general, written questionnaires have to be presented in the standard variety since dialects are usually spoken varieties and people are not used to reading and writing in their local dialect. However, a good alternative is to have some questions accommodated to a specific geographical distribution by inclusion of specific local dialect features such as subject-doubling, as illustrated in (5):

(5)	Misschien	ga'**k**	'et (e)	(**k**)	**ik**	wel	krijgen
	maybe	go-I	it	(I)	I	PART	receive
	Betekenis:	Misschien ga ik het wel krijgen					
	meaning:	'maybe, I'll get it'					

In the Netherlands and in Belgium the oral interviews in the SAND-project have taken place in the local dialect variety rather than in the standard variety or some regional Dutch or regional dialect variety, in order to avoid intermediate forms. Further, the conversations have been kept consistently in the local dialect in order to avoid intermediate forms between the local dialect and the standard language. Subsequently, we summon the assistance of another dialect speaker from the same community speaking the same local dialect in order to be able to interview the subject in his own local dialect. This assistant has the same social profile as the subject with respect to his age, socioeconomic background and autochthony.[16] This assistant is trained to translate orally all the questions into the local dialect variety and to administer these questions (which are tape-recorded) to the subject. In addition, the assistant is instructed to begin the interview session by having a conversation in the local dialect variety and to maintain this code so that no standard Dutch will interfere during these sessions. The assistance of another dialect speaker avoids the danger of accommodation (Giles and Powesland, 1975) to the speech of the 'strange' interviewer from outside the community or from a different sector of the community.

In the ASIS-project, the interviewer is often a speaker of the same local dialect, or has created the sentences with another collaborator/informant (see the previous section). This is the case, for instance, regarding all the Rhaetoromance inquiries that have been made on verb second, sentential particles related to sentence type and imperative clauses. As a result, all types of question are (when possible) translated on the spot into the dialect and not given in the standard language.

---

[16] Assistants are used as informants when there is disagreement in acceptability of certain judgments. They then make a note to the fact, apparently, variation is possible in these examples. In all other cases, the answers of the second informant are preferred because this elicitation is completely in the local dialect whereas the conversation/elicitation with the assistant is in a mixed repertoire, namely standard Dutch and dialect. Due to this mixed repertoire we consider the answers less reliable or more standard-like than the answers presented by the 'true' informant in interaction with the assistant.

3.5. Lexical items, context and pragmatic motivations

A striking task-effect of all grammaticality elicitation tasks is that a sentence may be judged to be unacceptable not because of its syntax, but rather simply because of the context, the lexical items chosen, and pragmatic considerations. Likewise, the use of non-natural intonation patterns or non-native phonemes can cause rejection to structures that transpire, in fact, to be perfectly legitimate. In the ASIS-project sentences are generally proposed within a context, especially when semantic differences are suspected. For instance, the contexts are very carefully prepared when testing phenomena related to sentence type: in imperatives and exclamatives, where the structures used can convey different meanings (for instance of advice, order, and prohibition for imperatives), the sentences tested are provided with a context compatible with only one possible interpretation. In the SAND-project, questions about negative concord and negative quantifiers, e.g. are accompanied with a very brief context on the basis of their mutuable nature with respect to semantics/pragmatics. More examples of this strategy can be found in the Swiss–German atlas in which each question was preceded by an everyday context, i.e. a little story, in order to create a discourse situation (cf. Bucheli and Glaser, 2002). It is very important to create a context which the subjects one is interviewing, perceive to be as neutral as possible.

Further complicating factors are that judgments may be based on estimated frequency of usage, or on the degree of semantic or pragmatic plausibility. Fig. 1 illustrates this. In 1995, a questionnaire was sent out in the Rhineland area which was offered entirely in standard German (cf. Cornips, 1996, Cornips and Corrigan, to appear). For each location, one native speaker of the local dialect completed the questionnaire. In this Rhineland questionnaire both the impersonal and adjunct middle with and without the reflexive *sich* were administered. For each variant (a), (b), (c), and (d), 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 their dialect.

The comments of the speaker in the Übersetzung phase in Fig. 1 reveals that he does not consider any of the middle constructions to be acceptable due to the fact that the lexical

		1 kommt vor/ ist manchmal zu hören		2 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.)			

Fig. 1. Part of a written questionnaire based on syntactic variants, e.g. the middle constructions.

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 as well as in the standard variety. Consequently, space for comments (as in Fig. 1) is necessary since they may 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. In oral inquiries, a number of researchers have conducted think-aloud tasks, in which subjects discuss their judgments as they are making them. The findings of think-aloud tasks thus far suggest that implicit knowledge is the primary basis for native-speaker decisions. Attempts to develop such tasks so that they are less subjective may be worthwhile avenues for future research (cf. Altenberg and Vago, 2002).

3.6. Translation

Most elicitation tasks inquiring about grammaticality judgments with respect to dialect varieties contain a written and oral 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 (especially in speech repertoires in which the standard variety is likely to influence the local dialect variety) is the repetition-effect; that is, the standard construction will simply be copied into the local dialect. With respect to written elicitation, the speaker has to write in dialect, although most informants, naturally, are not used to doing so. They, therefore, will concentrate more on the task of spelling and translating dialectal lexical elements that are not conventionally represented in the standard variety. As a result, more of their attention is focused on completing this insignificant aspect (from the researcher's perspective) of the task, when the researcher is actually more interested in the respondent's ability to handle syntactic variation. An alternative, for the purpose of evaluating the veracity of the subject's judgement is to make use of control sentences, for instance, to include some phenomena that do not exist in the dialects and to see how the speaker translates these sentences. Another alternative is to use so-called ''compliance tests'' (Greenbaum, 1973) in which the subjects have to transform a stimulus sentence in some way (e.g. by converting a question into a declarative sentence).

As we have discussed earlier, with respect to the northern Italian dialects, influence of the standard variety is very restricted. This can be seen by the speaker's refusal to translate ungrammatical sentences, or change them into a grammatical construction (i.e. the speaker avoids the standard structure). This is, for instance, the case with Aux to Comp constructions that do not exist in the northern Italian domain, but which are found in the standard variety. All the speakers convert these sentences into inflected embedded clauses. The same happens for complementizer deletion, where subjects systematically insert the complementizer when it is not present in the standard Italian version.

3.6.1. Other task-effects

Various task-effects have to be taken into account in any resultant analysis. Some task-effects can be minimised using one or more of the following strategies:

  (i) varying the order of the offered sentences since a habituation effect is likely to occur: that is, when a given sentence type is offered repeatedly, acceptability tends to increase;[17]
 (ii) using several tasks for the same syntactic features (cf. Bock, 1986);
(iii) asking the subjects for more examples of the same sentence type;
(iv) offering more alternative variants;
 (v) paying attention to the fact that subjects may give judgments on the basis of interpretability rather than grammaticality;
(vi) considering the distribution of grammatical versus ungrammatical sentences;
(vii) considering position of error in a sentence, the truth of the sentence, and sentence complexity (Altenberg and Vago, 2002).

Furthermore, the choice of the lexical elements could influence the 'positioning' of the speaker towards the standard or towards the dialect variety. Hence, it is always better to control for this variable and to use basic vocabulary.

These task-effects have to be taken into account, both in the design of elicitation methods and in the resulting analysis.

4. Difficult structures

Some structures are more difficult to elicit than others. For instance, in the ASIS-project it was found that some of the questions are very often misinterpreted. The same occurs with, for instance, a declarative clause with second person singular, which was extremely difficult to elicit since the subjects always translate these into imperatives or interrogatives. They do this since they are pragmatically coherent with second singular (cf. Oxford, 1982). In the SAND-project, questions with double negation appeared to be far too difficult to perform and showed a very high non-response. The subjects were asked to give the meaning of the construction in (6) that was uttered by the assistant:

(6)   Wim   denkt   dat   we   *nooit   niemand*   een   prijs   geven
        Wim   thinks   that   we   never   nobody   a   prize   give

The subjects could not make a motivated choice between the options—double negation of negative concord interpretation—offered in (7a) and (7b), respectively:

(7)   a   Wim denkt dat we *altijd iemand* een prijs geven.
             Wim thinks that we always somebody a prize give
        b   Wim denkt dat we *niemand ooit* een prijs geven.
             Wim thinks that we nobody ever a prize give

---

[17] Moreover, the relative order in which test sentences are presented to the subject has undoubtedly some influence on their judgments.

Syntactic elicitation provides no difficulties if structures are grammatical in the standard variety and ungrammatical in the dialects. In the ASIS-project, the subjects systematically refused those constructions, either by providing a grammatical alternative or by simply not being able to translate the sentence (see above the cases of Aux to Comp and complementizer deletion). However, the crucial point concerns structures which are optional, and only one of the two options is possible in the standard variety. Usually, subjects provide the same structure found in the standard and never the 'special' one of the dialect, thus preventing the researcher from finding it. These types of construction do not come out in the questionnaire and have to be proposed subsequently by the research outcomes, or they can be observed only through those subjects who have been trained to work as collaborators and who are, therefore, adept at systematically finding differences between the dialect and the standard. This issue is illustrated nicely by responses to aspectual reflexive *zich* (an optional element) in the Limburg dialects of The Netherlands, as illustrated in (8) (cf. Cornips, 1994):

(8) d'r   Jan   had   (*zich*)   in twieë minute e   beeke   gedrònke
     the   Jan   had   refl   in two minutes a   small beer   drunk

The construction with the reflexive *zich* is fully ungrammatical in the standard variety but acceptable in the local dialect. However, both constructions with and without the reflexive are acceptable in the local dialects. The written questionnaire shows that in only two out of 35 possible locations in the province of Limburg and its immediate surroundings, an answer with the reflexive is presented. In this case, the interference with the standard variety is so strong that the reflexive is not presented among the answers.

It appears that getting optional structures can only be done by listening to normal conversation or having a very good subject and asking them to give all the possibilities that come to mind. In order to do so, subjects have to be 'trained' to collaborate so that they come adroit at providing alternative structures.

5. Conclusion

In this paper, we have proposed several alternatives for standardising the design and implementation of grammaticality tasks and made various proposals regarding the manner in which non-linguists should be questioned about the (un)acceptability of syntactic features in their local dialects. We have discussed how grammaticality judgment tasks can be designed to enhance both their validity and their reliability. General problems with respect to the elicitation of syntactic features are, among others, (i) the use of implicit and explicit knowledge or prescriptive norms; (ii) the nature of speech repertoire of the dialect and the standard variety with respect to language of instruction; (iii) oral and written elicitation (or both); (iv) social variation as interfering factor; (v) the elicitation of optional constructions in the local dialect in cases where one option is ungrammatical in the standard variety and (vi) the misinterpretation of constructions for pragmatic reasons.

Acknowledgements

References

Altenberg, E.P., Vago, R.M., 2002. The role of grammaticality judgments in investigating first language attrition: a cross-disciplinary perspective. In: Talk Presented at the International Conference on First Language Attrition: Interdisciplinary Perspectives on Methodological Issues. Free University, Amsterdam, August 22–24, 2002. Ms Queens College and University of New York.

Auer, P. Europe's Sociolinguistic Unity, or: a Typology of European Dialect/standard Constellation. University of Freibourg.

Barbiers, S., Cornips, L., 2002. Introduction to syntactic microvariation. In: Barbiers, S., Cornips, L., van der Kleij, S. (Eds.), Syntactic Microvariation. Electronic publication of the Meertens Instituut. http://www.meertens.knaw.nl/projecten/sand/synmic/.

Bard, E.G., Robertson, D., Sorace, A., 1996. Magnitude estimation of linguistic acceptability. Language 72.1, 32–68.

Bock, J.K., 1986. Syntactic persistence in language production. Cognitive Psychology 18, 335–387.

Bucheli, C., Glaser, E., 2002. The Syntactic Atlas of Swiss German Dialects: empirical and methodological problems. In: Barbiers, S., Cornips, L., van der Kleij S. (Eds.). Syntactic Microvariation. Electronic publication of the Meertens Instituut. http://www.meertens.knaw.nl/projecten/sand/synmic/.

Carden, G., 1976. Syntactic and semantic data: replication results. Language in Society 5, 99–104.

Chomsky, N., 1986. Knowledge of Language: Its Nature, Origin and Use. Praeger, New York.

Cornips, L., 1994. Syntactische variatie in het Algemeen Nederlands van Heerlen (Syntactic Variation in Heerlen Dutch). Unpublished Dissertation. University of Amsterdam/IFOTT 6.

Cornips, L., Jongenburger, W., 2001. Elicitation techniques in a Dutch syntactic dialect atlas project. In: Broekhuis H., van der Wouden T. (Eds.), Linguistics in The Netherlands 2001, vol. 18. John Benjamins, Amsterdam/Philadelphia, pp. 57–69.

Cornips, L., Corrigan, K, to appear. Convergence and divergence in grammar. In: The Convergence and Divergence of Dialects in Contemporary Societies. In: Auer, P., Hinskens, F., Kerswill, P. (Eds.), Cambridge University Press, Cambridge.

Cowart, W., 1997. Experimental Syntax: Applying Objective Methods to Sentence Judgments. Sage Publications, Thousand Oaks.

Gervain, J., 2002. Linguistic Methodology and Microvariation in Language. Unpublished MA Thesis. University of Szeged.

Giles, H., Powesland, P.E., 1975. Speech Style and Social Evaluation. Academic press, New York.

Greenbaum, S., 1973. Informant elicitation of data on syntactic variation. Lingua 31, 201–212.

Labov, W., 1972. Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia.

Labov, W., 1996. When intuitions fail. Papers from the 32nd Regional Meeting of the Chicago Linguistics Society 32, 76–106.

Milroy, J., 2001. Language ideologies and the consequences of standardization. Journal of Sociolinguistics 5 (4), 530–555.

Muysken, P., 2000. Bilingual Speech. A typology of Code-mixing. Cambridge University Press, Cambridge.

Oxford, R.L., 1982. Technical issues in designing and conducting research on language skill attrition. In: Lambert, R. Freed, B.F. (Eds.), The Loss of Language Skills. Rowley, Massachusetts, pp. 119–137.

Pellegrini, G., 1982. Osservazioni di Sociolinguistica italiana. Italia Dialettale 45, 1–36.

Poletto, C., 2000. The Higher Functional Field: Evidence from Northern Italian Dialects. Oxford University Press, New York.

Rickford, J., 1987. The haves and have nots: sociolinguistic surveys and the assessment of speaker competence. Language in Society 16-2, 149–177.

Sanga, G., 1991. I metodi della ricerca sul campo. Rivista italiana di dialettologia 15, 165–183.

Schütze, C.T., 1996. The Empirical Base of Linguistics. Grammaticality Judgments and Linguistic Methodology. The University of Chicago Press, Chicago.

Vogelaer, G., De Annemie N., Guido, V.W., 2002. Complementizer Agreement in the Flemish Dialects. In: Barbiers, S., Cornips, L., van der Kleij S. (Eds.), Syntactic Microvariation. Electronic Publication of the Meertens Instituut. http://www.meertens.knaw.nl/projecten/sand/synmic/.

Weiß, H., 2002. Three types of negation: a case study in Bavarian. In: Barbiers, S., Cornips, L., van der Kleij S. (Eds.), Syntactic Microvariation. Electronic Publication of the Meertens Instituut. http://www.meertens.knaw.nl/projecten/sand/synmic/.