
Compiling a Dictionary of an Unwritten Language: A Non-corpus-based Approach

Jacques van Keymeulen, *Department of Dutch Linguistics, Ghent University, Belgium* (jacques.vankeymeulen@ugent.be)

Abstract: In this article an account is given of the experience in fieldwork by the Dictionary of the Flemish Dialects (*Woordenboek van de Vlaamse Dialecten, WVD*), Ghent University, Belgium. The focus is on the practical aspects with regard to methods of lexicographic fieldwork. It is maintained that the analysis of 'metalinguistical conversations' with groups of respondents in which their lexicographic competence is explored, is a suitable way of collecting lexicographic data. Fieldwork by correspondence (questionnaires) can amplify and verify the data collected through interviews.

Keywords: LEXICOGRAPHY, UNWRITTEN LANGUAGE, DIALECT, REGIONAL DICTIONARY, FIELDWORK, GENERAL VOCABULARY, DUTCH, SOUTHERN DUTCH, FLEMISH, BRABANT DIALECT, LIMBURG DIALECT, THE NETHERLANDS, BELGIUM, SYSTEMATIC ARRANGEMENT, METHODOLOGY, QUESTIONNAIRE, INTERVIEW, WORD ATLAS, LANGUAGE VARIATION

Samenvatting: Een woordenboek samenstellen van een ongeschreven taal: een niet-corpusgebaseerde benadering. In dit artikel wordt een verslag gegeven van de ervaringen met veldwerk van het *Woordenboek van de Vlaamse Dialecten (WVD)*, Universiteit Gent, België. De klemtoon ligt op de praktische aspecten met betrekking tot lexicografische veldwerkmethodes. Er wordt verdedigd dat de analyse van 'metalinguïstische conversaties' met informantengroepen, waarbij de lexicale competentie wordt geëxploreerd, een goede manier is om lexicografische gegevens te verzamelen. Veldwerk per correspondentie (vragenlijsten) kan de gegevens die mondeling verzameld zijn, vervolledigen en controleren.

Sleutelwoorden: LEXICOGRAFIE, ONGESCHREVEN TAAL, DIALECT, REGIONAAL WOORDENBOEK, VELDWERK, ALGEMENE WOORDENSCHAT, NEDERLANDS, ZUID-NEDERLANDS, VLAAMS, BRABANTS, LIMBURGS, NEDERLAND, BELGIË, SYSTEMATISCHE ORDENING, METHODOLOGIE, VRAGENLIJST, INTERVIEW, WOORDATLAS, TAAL-VARIATIE

Introduction

Unwritten languages exist everywhere. Many standard languages in Europe span a wide variety of purely oral — unwritten — dialects. The 'traditional'

dialects are not regional 'accents' of a standard language, but are ancient language varieties that for historical, political and sociological reasons did not reach the status of a 'cultural language'; they are linguistically characterized by the presence of geographical differentiation. They are sometimes not mutually intelligible and do not have, with a few exceptions, a tradition of written texts.

The aim of this article is to share the experiences with regard to the methods of collecting dialect words/dialect meanings used by the editorial board of the Dictionary of the Flemish Dialects (*Woordenboek van de Vlaamse Dialecten*, WVD) at Ghent University, Belgium. The WVD is also counselling local dictionary projects, for which a manual has been written (Van Keymeulen 2003a). We hope that some of our experiences will be of interest to other projects aiming at making dictionaries for oral languages. The focus of the article is on the practical aspects of the methodology of data collection, more particularly on the fieldwork to be performed (par. 1). We will briefly comment on the presentation of the macro- and microstructure of the WVD and its two cognate projects (par. 2). In par. 3 we present the conclusions.

There is a vast literature on the methodology of fieldwork, especially for sciences such as anthropology, ethnology, sociology and oral history. For a theoretical background with regard to participant observation, interview techniques, questionnaires, etc., we refer the reader to the relevant handbooks. In what follows, we restrict ourselves to the methods used by the WVD, the result of 30 years of experience.

Flemish dialects

Flemish dialects are spoken in the south-western corner of the Dutch language area, i.e. two provinces of Dutch-speaking Belgium (West- and East-Flanders), plus two smaller areas in France (French-Flanders) and the Netherlands (Zealand-Flanders). The traditional Flemish dialect vocabulary is rapidly disappearing under the pressure of the Dutch standard language, the official language of the Flemings. In France, where the dialect is not 'roofed' by Dutch, but by French, its disappearance is imminent. It should be noted that the term 'Flemish' is used here in its restricted — dialectological — sense. In popular speech, both Flemings and foreigners use the term to denote the 'Belgian' accent (and other particularities) of the Dutch standard language.

Dictionaries of the Southern Dutch Dialects

The set-up of the WVD, which started in 1972, is largely parallel with that of two cognate and older projects: the Dictionary of the Brabant Dialects (*Woordenboek van de Brabantse Dialecten*, WBD) and the Dictionary of the Limburg Dialects (*Woordenboek van de Limburgse Dialecten*, WLD). Both dictionaries started in 1960 at the Catholic University of Nijmegen in the Netherlands; from 1991 onwards, an additional editorial board for WBD and WLD has been estab-

lished at the Catholic University of Louvain in Belgium. The three dictionaries together cover the southern Dutch language area.

The methodology of the dictionaries, which was initially developed by prof. A.A. Weijnen (at the Catholic University of Nijmegen), has already been described in Kruijssen en Van Keymeulen (1997), so I limit myself to a few major points. All three dictionaries are arranged systematically (with alphabetical indexes) and combine a dictionary with a word atlas, since for every dictionary the focus is not on one dialect but on a group of related dialects. The dictionaries are divided into three separate parts: I. Agricultural vocabulary; II. Technical vocabularies; III. General vocabulary (i.e. the vocabulary not restricted to professional activities). Every fascicle of the dictionary is devoted to a certain conceptual field (e.g. 'housing', 'wild animals', etc.) and consists of a series of concepts related to it (onomasiological arrangement). For every concept the heteronymy (the different lexemes which can be used to refer to the concept) in the different dialects of the area under investigation is presented, together with general indications as to frequency and location (details with regard to phonetics and location are kept in an automated database). (For further information, see Kruijssen and Van Keymeulen 1997 and the website of the three projects at <<http://www.flwi.ugent.be/dialect>>.)

The aim of the three dictionaries is to document the vocabulary of the oldest layer of the traditional dialects and its geographical patterns. Since the traditional dialects are not written, the bulk of the data is taken from interviews or consists of answers to questionnaires filled in by hundreds of volunteering elderly dialect speakers. To these data, material taken from older investigations or from written sources such as older dictionaries is added.

The metalexicographical questions with regard to aim ('Why?') and target users ('For whom?') are answered as follows. The ultimate aim is to document a disappearing vocabulary and its lexico-geographical patterns for scientific reasons. This geographical aim explains the neglect of collocations (expressions, proverbs, etc.). The dictionaries are strictly descriptive with the future scientist as target user. Practical usage of the dictionaries is considered marginal. Although the dictionaries enjoy a wide popular support today, in future the scientific user will come more to the fore as the dialect vocabulary becomes extinct. The dictionaries provide data for the future study of the lexical history of the Dutch language; they are complementary to the dictionaries of both the historical and present-day periods of Dutch, which are based on written text corpora. The metalanguage of the three dictionaries is standard Dutch.

In what follows, we will not dwell on all the possible answers to the questions 'Why?' and 'For whom?', which determine both the macro- and micro-structure of any dictionary, since they may vary widely according to the different sociological and linguistic contexts in which a dictionary has to function. In case a purely oral language has to be turned into an official and written language, a 'norm' has to be chosen and codified and an orthography has to be agreed upon. The problems with regard to codifying a dominant dialect, of

which all the other dialects are to be considered as 'variants', may engender heated discussions. Solutions to questions pertaining to codification are in fact too important to be left to lexicographers. It is the political authorities and the language policy makers/financing bodies that should provide the metalexicographical answers in this respect. The introduction of standard Dutch in the north of Belgium, incidentally, provides good examples of linguistic frustration, linguistic struggle, linguistic planning, etc. (see Van Keymeulen 2003b).

1. Data collection

By definition, it is hardly possible to make a dictionary of an *oral* language on the basis of a text corpus. For some 'unwritten' languages or language varieties texts are available, but mostly they are scarce, if not absent altogether. Although it is certainly worthwhile always to collect all the available texts, a corpus-based approach is not the best way to success, given the fact that a collection of texts will at best form an unbalanced and unrepresentative corpus. Any bit of lexical information, however, taken from whatever type of source, may prove interesting when it comes to preparing the fieldwork. As we shall see later, fieldwork needs an input in order to maximize its success.

In what follows, we focus on the different stages of the lexicographic fieldwork and on the way the five key questions pertaining to data collection may be answered. These questions are: 'What?', 'Who?', 'Where?', 'How?', and 'How much?'.

1.1 What?

The question 'What to collect?' pertains to the macro- and microstructure of the dictionary: what words/collocations are to be collected, and what information about the words has to be gathered? The answers to the 'What?' are very central ones, which should, after careful reflection, be answered in the very first stage of the project.

Oral language traditions may be 'roofed' by a linguistically related standard language, as is the case for all dialects both in the Netherlands and Flanders. This has the effect that the non-standard lexicon coincides partly with the lexicon of the standard language: some differences are restricted to phonology, while others are lexical proper. In a case like this, one has to decide between a confrontative and a contrastive dictionary (the terms are Wiegand's). In a confrontative dictionary, the aim is to present the totality of the lexicon of the non-standard language; in a contrastive dictionary the focus is on the lexical and semantic differences (neglecting the purely phonological ones) between dialect and standard language.

Every language varies both geographically and socially (age, class, level of education and gender being the typical parameters). A clear position has to be taken with regard to the inclusion of the horizontal (geographical) and vertical

(social) variation. Another important decision pertains to the inclusion or exclusion of terminologies of professions or other socially restricted domains. Some occupations, e.g. farming, may be so omnipresent in the culture that the vocabulary accompanying it can be considered essential for the language as a whole, whereas other, more restricted, terminologies, e.g. the vocabulary of the tailor or the smith, may be excluded. In principle, all variables with regard to linguistic variation should be kept uniform, and if not, the words/meanings should be labelled in the dictionary.

Finally, decisions have to be made with regard to the microstructure: What elements of e.g. the phonology, morphology and syntactical valency of the words have to be collected? What labels should be introduced?

Generally speaking, we advise the lexicographer not to undertake more than can be handled. It is better to succeed in reaching a modest aim, than to fail trying to adapt an unmanageable mass of data. Most local dialect dictionaries in Flanders, for instance, are of the contrastive type, and restrict themselves to the 'general vocabulary' (excluding dialectal terminologies) of the traditional dialect of the oldest generation. Their microstructure mostly contains only pronunciation, meaning, collocations and example sentences. The three regional dialect dictionaries mentioned above arrange concepts, together with the accompanying dialectal heteronymy in the area of investigation. Only the location and pronunciation of the lexical items are accounted for.

The answers to the question 'What?' determine the answers to the question 'Who?'

1.2 Who?

The answers to the question 'What?' determine the answers to the question 'Who?'

In the absence of a corpus of written texts, the word material for dictionaries of oral languages has to be collected by way of fieldwork. The quality of the data depends heavily on the choice of the volunteering respondents, the profile of whom is determined by the answers to the question 'What?'. Two types of volunteers may be distinguished: intermediaries and dialect respondents proper. The task of the intermediary is to look for suitable dialect respondents (for a specific subject) and to facilitate the fieldwork by introducing the researcher to potential collaborators within the local community.

An intermediary should meet the following requirements:

- take an interest in the dictionary project;
- be highly respected in the local community; and
- know a large number of people at all levels of society.

In practice, intermediaries can be found amongst the clergy, (school)teachers, school directors, chairpersons of organisations, etc. It is these people who have

the necessary educational background to understand the importance of the dictionary project and who may be willing to introduce the researcher to the local community. In the case of the WVD, a campaign directed towards parish priests to find respondents for agrarian vocabulary proved to be highly productive.

A good relationship with the intermediary is very important since one has regularly to revert to him/her. Lexical knowledge, as it happens, is distributed unevenly in the language community, and one has to recruit different respondents for different topics: farmers know more about the words for wild plants than other people; women know more about the kitchen than men, etc. The WVD regularly convenes with the intermediaries/respondents in order to keep them motivated by communicating interim results and by organizing social activities.

The profile of the ideal respondent should be established very carefully, taking into account the options with regard to the macrostructure of the dictionary. A respondent should meet a list of both subjective and objective requirements.

The subjective requirements amount to:

- an interest in the dictionary project;
- intelligence;
- willingness;
- communicability; and
- suitability for the fieldwork according to his/her own opinion.

The objective requirements usually have to do with sociological and geographical parameters, and with the subject matter under investigation. In the case of a dictionary of a traditional local dialect in danger of disappearing under the pressure of a standard language, the objective requirements to be met by the respondents are:

- being of an advanced age;
- being of a lower social class;
- having a low level of schooling;
- having been locality-bound, i.e.
 - having grown up/lived all his/her life in the same locality,
 - having exercised an occupation/a trade in that same locality,
 - having both parents come from that same locality, and
 - being married to someone of that same locality;
- being expert in the topic under investigation.

The question whether men or women are more suitable as respondents in general is a tricky one. It has been observed regularly that women are more sensitive to language change than men, hence they tend to be more influenced by standard varieties. On the other hand, they are often more home-bound, which makes them linguistically more conservative. In practice, many dictionaries of traditional local dialects rely on Trudgill's NORM-respondent, the Non-educated Old Rural Male.

In some cases, a conflict may arise between the subjective and the objective requirements. The editorial board of the WVD experienced the paradox that some very suitable dialect speakers were unfit to cooperate with the project because they could not understand its purpose.

Respondents should be tested. In the case of the WVD, the statements of respondents who could not spontaneously produce the dialect word for 'butterfly' or 'hedgehog', but used the 'dialectized' standard Dutch lexeme instead, were regarded as doubtful. A further requirement is that the biographical data of the respondent and other relevant information pertaining to his profile should be noted down.

1.3 Where?

The geographical scope of a dictionary may vary from a single locality or a region to a whole language community. If the aim of the dictionary is to codify a vocabulary, a choice may have to be made as to the socio-economical dominant region. Regional dictionaries of the WVD-type only aim to describe a geographically differentiated dialect landscape. The word maps, however, indicated that the Brabant dialect is dominant in northern Belgium. Brabantine words turned up in both Flemish and Limburg dialects; the opposite was never the case.

The editorial board of the WVD could rely on a long tradition of dialect investigation and a fairly good knowledge of the lexical geography of the dialects. Although the heteronymy for every concept has its own geographical pattern (the pattern of the dialect words for 'butterfly', for instance, is totally different from the pattern of the words for 'duck'), it was possible, on the basis of the comparison of many hundreds of word maps, to detect recurrent patterns. A number of 'lexical areas' could be distinguished, i.e. areas which regularly have their 'own' word for a given concept. In every lexical area, one rural 'reference point' for an in-depth interview with a group of respondents was selected. To these, six urban points were added to account for the major urban dialects. A number of interviews were conducted by a researcher to obtain sound (orally collected!) lexical information for every lexical area (the minimal goal). In a later phase, questionnaires were sent out to respondents in as many localities as possible to make it possible to draw word maps and to detect the course of the isolexes (the maximal goal). Since one has to depend on volunteers, the maximal goal is hard to meet.

1.4 How?

The lexicon of an unwritten language can be collected by several methods:

- self-observation;
- observation of spontaneous speech;
- recording and transcribing 'free speech';
- purposive systematic fieldwork, i.e.
 - recording and transcribing 'free speech',
 - oral investigation (interviews/questionnaires), and
 - investigation by correspondence (questionnaires).

Many dictionary projects derive their lexical information from a variety of sources. Some methods are complementary; the results of one method may be used as the input for another. The methods differ firstly with respect to the relative validity of the data: lexemes taken down from spontaneous speech are thought to be of higher quality than those collected purposively. Secondly, the methods differ with regard to the possibility of being carried out systematically. Thirdly, methods may vary according to the type of vocabulary one wants to collect. Some elements of the lexicon are more easily collected by purposive questioning than others, because the introspective capacity of a language user differs according to the different lexical types. The main difference in this respect is between the 'open' (e.g. substantives, adjectives, verbs) and the 'closed' word categories (e.g. abverbs, prepositions, conjunctions). A classic dichotomy, finally, is the distinction between direct and indirect methods, i.e. fieldwork conducted orally or by correspondence respectively. In our view, it is the presence of an intermediary, rather than the medium of communication, that makes a method direct or indirect.

In what follows, we will briefly discuss the various methods of data collection and point out the advantages and disadvantages of each of them.

1.4.1 Self-observation

The advantage of writing a dictionary of one's own language is self-evident. A word collection can indeed be expanded greatly when using one's own language competence. There is, however, a risk. A lexicographer may be influenced by the standard variety of the language — more than he is aware. A phonologically adapted borrowing from a standard variety, for instance, may have replaced the traditional dialect word in his linguistic memory. The question is whether the lexicographer's profile as 'respondent' is in accordance with metalexicographical considerations. However, data drawn from one's own language memory may always serve as an input for other methods.

The above-mentioned risk firstly pertains to word collecting. The lexicographer's language competence is of course invaluable when it comes to describing the meaning of the lexemes or adding other microstructural elements.

1.4.2 Observation of spontaneous speech

The word material can be extended by noting down overheard words or expressions. Many a lexicographer of a local dialect dictionary carries a notebook for this purpose. Since the word material is collected in real language situations, the advantage of this method is the high validity of the data. This advantage is even more important for the closed word categories, which are harder than the open ones to inventory by explicit questioning. The disadvantages of the method pertain to the unknown sociological profile of the speakers and the unsystematic nature of the method. It is, however, a good way to collect material for further fieldwork and it certainly is a good method to collect real and lively example sentences. In many dictionaries example sentences sometimes appear too 'construed', and sentences taken from spontaneous speech often add realism and liveliness to a dictionary.

1.4.3 Recording and transcribing 'free speech'

In an attempt to overcome the problem of the availability of few or no written texts, some lexicographers have tried creating the necessary texts themselves by making transcriptions of interviews with respondents. The method amounts to choosing a good respondent and making him/her talk as much as possible about a specific subject. The interview is then fully transcribed. On the basis of the written transcriptions of the language material provided by the respondent, corpus analysis procedures may be carried out.

Although the validity of the lexicographic data is very high, since it is derived from spontaneous speech, the method is seldom used because it is very time-consuming. Moreover, it is hard to collect a vocabulary systematically by this method only, unless the subject matter of the interviews is restricted and the interviews themselves are highly structured.

In the case of the WVD, the existing transcriptions of the 400 odd recordings of 'free' dialect speech collected within the framework of a project on dialect syntax, are as yet not used for lexicographic purposes, since there are more efficient ways for collecting vocabulary. For the open word categories, the advantages of the method were deemed not to counterbalance the effort. For the words of the closed categories, however, which are highly frequent and harder to collect by way of questions, the collection of transcriptions may be used as corpus.

In the next paragraph, we will present a fieldwork procedure which includes relatively unstructured interviews. Through this process, a collection

of recordings which may afterwards be used for a variety of purposes, can be made.

1.4.4 Systematic fieldwork: in search for concepts, words and meanings

In the absence of written texts, fieldwork is the only systematic way of collecting material for a dictionary of an unwritten language or language variety. In what follows, we will discuss a number of fieldwork procedures which may be of use both for a small-sized local dictionary and for a comprehensive geographically-oriented dictionary. We assume that little or nothing is known about the lexicon under investigation and we consider all data collected by other methods, as an input for more purposive actions.

The methods we have discussed so far, all share the characteristic that the vocabulary is collected unsystematically. None of these guarantees completeness. In what follows, we describe the fieldwork procedure carried out by the editorial board of the WVD. It aims at systematically collecting the lexicon of an unwritten language, i.e. exploring the vocabulary topic by topic. We will first pay attention to the 'classification of reality' (par. 1.4.4.1) and then proceed to the successive stages of the fieldwork procedure: preparation (par. 1.4.4.2), group interviews (par.1.4.4.3), and large-scale investigation by correspondence (par. 1.4.4.4.). In a separate paragraph (par. 1.4.4.5.), we discuss a number of question types.

1.4.4.1 A classification of 'reality'

The fieldwork is conducted systematically, i.e. subject by subject. For this purpose, human experience and the surrounding world have to be classified in a coherent system. It has to be stressed that this system is not a linguistic but an ontologic one, i.e. it is a classification of reality, not of the lexicon. The system only serves as framework for a thematic lexicographic investigation.

The classification used by the three editorial boards of the regional dialect dictionaries of southern Dutch (WBD, WLD, WVD) for the collection of the 'General vocabulary' is based on the 'classical' classification of Hallig-Von Wartburg (1952), adapted by Frissen (1981) and Van Keymeulen (1992), who was inspired by the classification of the WALD. The classification tries to account for what Weijnen and Van Bakel (1967: 40) called 'the concrete coherence of things in daily life' as experienced by the dialect-speaking community. A detailed discussion of the principles according to which the classification below has been made is not possible here. Generally speaking, man is placed at the centre of things, and reality is assigned to him in ever broadening circles: domestic life, society and the surrounding world. Departing from the four main sections, the classification forms an ever finer hierarchy of sections, indicating a particular subject matter (e.g. 4.2 'Plants and animals' > 4.2.2. 'Animals' > 4.2.2.3 'Reptiles and amphibians' > 4.2.2.3.2. 'Amphibians').

Classification

1. Man
 - 1.1 Man as a physical being (parts of the body, movements, ...)
 - 1.2 Clothing
 - 1.3 Internal reality of man (emotions, character, ...)
2. Domestic life
 - 2.1 The house
 - 2.2 Family life/phases of life (birth, childhood, marriage, ...)
 - 2.3 Eating and drinking
3. Society
 - 3.1 Social behaviour
 - 3.2 School
 - 3.3 Religion
 - 3.4 Amusement, festivities, art
4. The world around us
 - 4.1 The material world (weather, ...)
 - 4.2 Plants and animals
 - 4.3 The abstract world (time, space, form, ...)

Since the classification is made with 'the concrete coherence of things in daily life' of the dialect-speaking community in mind, is it a classification of the *culture* rather than the lexicon of that community. The main framework of the classification has universal value, and adaptations with regard to specific cultural contexts are only needed at its lower levels.

In the case of the three dictionaries of the southern Dutch dialects, the classification has been adapted during fieldwork and has consequently been used for the presentation of the macrostructure, which is issued in a series of thematic fascicles, each of which is devoted to a specific subject. In our view, however, a systematic (theme by theme) exploration of the lexicon is also recommendable for alphabetically arranged dictionaries.

1.4.4.2 Preparing for the fieldwork

Preparing for the lexicographic fieldwork has two aspects: (a) studying the subject matter itself; and (b) collecting all lexical data already available concerning the subject matter.

In many cases, good descriptions of the daily-life culture of the language community under investigation are absent, and the lexicographer has to do the necessary ethnological and anthropological research partly or even wholly by himself (see par. 1.4.4.3). In the case of the WVD, a subject such as 'birds' proved to be relatively easy in this respect, since good ornithological handbooks are available. For other subjects, especially the traditional crafts and

trades (such as cooper, basket maker, thatcher, etc.), the documentation gathered by the editorial board turned out to be the most extensive existent — often to the amazement of folklorists. An additional advantage of the systematic fieldwork method is that cooperation between ethnologists and lexicographers is possible and in fact desirable.

During the preparatory phase, the lexicographer should also collect the already available lexical data in written sources (texts if any, already existing dictionaries, scientific studies, etc.), which can afterwards be used as input for the interviews. In the preparatory phase there is also the opportunity to collect items which may stimulate the memory of the respondents (e.g. photographs, drawings, etc.).

The main aim of the preparatory phase which ideally can be restricted to reading specialist (ethnological) literature, is to gather information for structuring the interviews for the fieldwork proper.

1.4.4.3 The group interview: in search of quality

When the preparatory phase is completed, one can turn to the fieldwork proper: the in-depth interviews with respondents. The interviews may be both ethnological and lexical in nature, depending on the amount of ethnological literature one was able to find.

It should be emphasized that from a scientific point of view the source of lexical/ethnological information in fieldwork is *not* the respondent, but the interview itself, which in its oral or written form is the result of an interaction between interviewer and interviewee. An interview is an 'artefact'. The fieldworker should be aware that the lexical knowledge in the mind of the respondent may very well be stored and organised quite differently from what has been expected. Cicourel (1988: 907) warns us against 'questions that may not evoke an appropriate recall because of being phrased differently from the storage format of the informants' or respondents' experiences and their use of semantic memory'.

An efficient elicitation of lexical data (words or meanings) in metalinguistic discourse presupposes the establishment of a 'common ground' for meeting between interviewer and interviewee. This common ground should include a common understanding of the aims of the interview and a sound knowledge of the conceptions and motivations of the respondent. The advice of Cicourel (1988: 910) should be taken to heart: 'Elicitation procedures ... should seek to maximize the activation of schemata or mental models that have been constructed by informants.'

In the first phase of the lexicographic fieldwork concerning a relatively unknown culture/language variety, the aim of the procedure is threefold: (a) the establishment of an inventory of concepts; (b) the establishment of an inventory of lexically relevant concepts; and (c) the establishment of the way the meaning of the lexemes is stored in the semantic memory of the respondents.

The first aim pertains to an analysis of the culture of the language community and its context, apart from linguistic considerations. In doing this, the lexicographer may be obliged to enter the field of other sciences, anthropology, ethnology, oral history, etc., depending on the level of the already existing knowledge about the culture of the language community. If this knowledge is insufficient, and if good descriptions of the culture are unavailable, the lexicographer may benefit from the existing handbooks on fieldwork in those sciences.

When an inventory of concepts (things, actions, attitudes, etc.) is completed, i.e. when there is enough *encyclopedic information* available, the fieldworker can proceed to the second aim: to try and obtain an insight into the structure of the semantic field related to the subject matter under investigation. In the case of 'birds', for instance, this in practice means an investigation into folk taxonomy, which may differ widely from scientific taxonomies in biology. Different kinds of birds may be denoted by the same lexeme; some species may not be named at all. In practice, the procedure amounts to investigating how the concepts are structured and what concepts are named.

The third aim is that the interview reveal the way meanings are stored in the semantic memory of the respondent. The meaning of substantives for concrete objects (e.g. *table*) is explained in a way different from an explanation of the meaning of an adjective denoting a character trait (e.g. *miserly*). There is a good chance that the respondent will describe the meaning of *table* by pointing to the function of the object, whereas *miserly* could be explained by giving an example of an action by a miser. Such semantic indications are very useful for framing explicit questions in the later stages of the fieldwork, and for the description of the meaning in the dictionary afterwards. Respondents usually describe the meaning of a word by using synonyms and antonyms, translations into another language, analytic definitions with semantic components, showing the thing or pointing to it, drawings, encyclopedic descriptions, prototypical examples, and descriptions of situations in which the word could be used.

For the sake of clarity, the three aims above have been presented as if they could be pursued consecutively. This is, however, seldom possible. A metalinguistic conversation derives its complexity from the fact that the interview constantly provides information of all three types. It is not an easy task to ensure that all necessary information, both encyclopedic and lexical, is elicited from the respondent when conducting the conversation.

In this phase, the actual fieldwork of the WVD is carried out by a researcher by way of a number of semi-structured interviews ('guided conversations') in the different reference points with groups of four to six respondents assembled by an intermediary, all resident in the same locality. The subject matter (e.g. 'birds') is divided into subthemes (on the basis of the above-mentioned classification) which are subsequently used to structure the conversation (e.g. concerning seabirds, birds of prey, forest birds, etc.). The respondents are invited to talk freely about the subthemes. In the course of the conversation, the

data from written sources collected in the preparatory phase are introduced by the fieldworker. In the case of the subject 'birds', for instance, the respondent groups were asked to comment on a series of pictures taken from a handbook on ornithology, and at the same time they were invited to describe the meaning of a number of words already known.

There is a wide variety of interview techniques, each of which has its own merits and limitations, and a multiplicity of uses (see Fontana and Frey 1994). Every technique, however, should take into account the cultural context in which the interview has to be conducted (e.g. politeness rules). It goes without saying that an interviewer should have the necessary personal qualities, of which flexibility and empathy are not the least important.

In our experience, the group interview is a very productive method to elicit ethnological/lexicographic data. Labov's 'observer's paradox' can be overcome, especially when the group is visited regularly. Interviewing a group has the following advantages:

- the respondents feel self-confident, because the researcher is in the minority;
- the respondents are used to speak the specific language variety with each other, hence the tendency to adapt to the language of the researcher is low; and
- the respondents can correct each other, therefore the data are very comprehensive.

Merton et al. (1956) note three specific skills needed by a group interviewer:

- he has to keep one person or a few persons from dominating the group;
- he should encourage reluctant respondents to participate; and
- he must obtain responses from the entire group to ensure the fullest possible coverage of the topic.

A group interview may yield data of a very high quality, provided the interviewer was able to manage the group dynamics efficiently. König (1982/1983: 477-478) distinguishes 'primary' and 'spontaneous' material in oral investigations. Primary material consists of the answers to questions; spontaneous material is the unintentional information given by a respondent when talking. The (partly) unstructured nature of the group interview guarantees that the fieldworker will acquire yet unknown information.

In case of a long-term dictionary project, the participants of the group, being interviewed regularly, obtain in-job training during the process and develop from mere 'respondents' into 'language counsellors', who eventually may even comment on proof versions of the dictionary text.

The WVD tries and conducts group interviews in each of the lexical areas. The data yielded by one group is used as an input for the next interview in another area. In the course of the fieldwork, the interviews become ever more structured, till it is possible to frame questions which can be used for fieldwork by correspondence.

Although the group interview should be loosely structured during this phase of the fieldwork, it is clear that targeted questions should be asked. In par. 1.4.4.5, we present a list of question types which can be used either in interviews or in questionnaires.

1.4.4.4 In search of quantity: questionnaires

On the basis of the results of the group interview(s), the lexicographer is in a position to prepare good questionnaires for large-scale investigation, and move to a more standardized form of research, in which comparable data are produced. Large-scale investigations are inevitable if one wants to document the geographical patterns in a lexicon.

In the case of the WVD, the intermediary takes the place of the researcher and conducts a highly structured interview with a respondent by means of a questionnaire. This is however an ideal situation; in many cases intermediary and respondent are the same person. Research by correspondence has the obvious disadvantage that the researcher himself is not present and that the situational setting of filling in the questionnaire cannot be supervised. Filling in questionnaires of course presupposes that the language has a written form, and that the intermediaries/respondents are capable of using it. Since the Flemish dialects are linguistically related to the Dutch standard, dialect words can be written down in a 'Dutchified' form (see par. 2.2 below).

1.4.4.5 Types of questions (interviews/questionnaires)

In lexicographic interviews/questionnaires, apart from the classical dichotomy between 'open' and 'closed' questions, five main types of questions can be distinguished: exploratory questions, encyclopedic questions, onomasiological questions, semasiological questions and task questions. In 'open' questions, the respondent is entirely free to answer as he pleases. 'Closed' questions, where the respondent has to choose between yes or no, or has to pick an answer out of a series of suggestions, are not frequently used in lexicographic fieldwork. (For a theoretical background on questioning and further reading see Atteslander 1988.)

Exploratory questions

Exploratory questions focus on the external world or on language; usually a certain subject is introduced and the respondent is invited to give as much

information as possible. This type of question is typical of the initial phase of fieldwork, and should be used in the group interview. In questionnaires, the technique is not very productive.

There are techniques which aim to stimulate the 'word-finding' process. When a certain subject is introduced, a number of questions may be asked (e.g. 'What is the opposite?' 'Are there different kinds of it?') in order to expand the word collection.

Encyclopedic questions

Encyclopedic questions in principle deal with the extra-linguistic world; these questions are not aimed at language but the answers will of course contain lexemes. Encyclopedic information should be gathered during the group interview. Lexicographic questions in questionnaires, however, may be masked as encyclopedic ones, as in: 'How do you feel when everything seems to turn round?'. A question like this leads to the answer 'dizzy'. Provided the question is well framed and unequivocal, it is a useful technique to avoid a complex onomasiological definition.

Onomasiological questions

Onomasiological questions describe/define/frame a concept and ask for a lexical expression. The technique is widely used for the open word classes. The definition of an onomasiological question can take a variety of forms: pictures, analytical definitions using semantic/encyclopedic components (e.g. the description of a prototypical example), or framing a context in the object language in which a lexical expression should be used (e.g. by way of sentences in which words should be filled in). The question-frame should lead the respondent unequivocally to the desired answer. The choice of the definition is dependent on the type of lexical item under investigation. Lexemes for concreta can be elicited by using analytical definitions of the concept; the more abstract or grammatical the meaning, the more one has to resort to other definition types, which try to imitate object language.

An old-fashioned onomasiological elicitation technique in Flemish dialect lexicography was to define a concept by one standard Dutch word. The respondent was expected to translate the word into his dialect. This technique, however, proved to be problematic since many respondents merely transposed the Dutch word phonologically in dialect, instead of translating it.

The question whether answers (i.e. lexemes) can be suggested, is still a matter of debate. This technique, which aims at stimulating the memory, is dangerous because it may provoke echo-answers. An all too cooperative respondent may select a lexeme he knows, but does not use — it is difficult to tell the two apart. The WVD, however, has been forced to suggest answers because for some subjects the traditional dialect words were no longer in active usage.

It proved that the oldest words could only still be recognized by the respondents; they were unable to remember them spontaneously any more. Suggestions can be presented in an open or a closed series; in the former, the respondent may add yet another word, in the latter, he has to select an answer.

Semasiological questions

In semasiological questions, a lexical expression is presented, and a 'meaning' (conceptual meaning, connotation, register or use) is asked for. From the respondent's point of view this technique is more demanding than the onomasiological one, where just one word may suffice in many cases. Semasiological questions require more thinking and noting down, which sometimes result in all too short or ill-conceived answers. This type of question calls for a good and reliable respondent.

Semasiological questionnaires presuppose onomasiological ones (one has to collect the lexemes first) and are often used to verify the semantic anticipations underlying onomasiological questions. The aim is to collect data about the meaning of a lexeme. The semantic information given by respondents may take a variety of shapes, which have to be interpreted by the researcher and translated afterwards into 'scientific' lexicographic definitions.

Suggestions as to the meaning or the usage of a lexeme, in an open or closed way, may streamline the answers.

Task questions

The aim of task questions is to collect words in context: collocations and sentences. The respondent is invited to produce expressions, proverbs or example sentences in which a given word is used. The dictionary of the dialects of Drenthe (a province in the north-east of the Netherlands), for instance, was compiled largely on the basis of a collection of example sentences produced by respondents. Cards with words on them were sent out to a large number of respondents, who were asked to illustrate the meaning of each word by constructing a sentence with it.

Cyclicity and feedback

Lexicographic fieldwork carried out in a relatively unknown language area should take advantage of the possibility of using cyclical questioning procedures. The answers to onomasiological questions can be used as an input for semasiological ones, which may provide new words and meanings. The alternation of onomasiological and semasiological points of view thus creates a cyclical questioning procedure until all words and meanings within a given conceptual field are accounted for.

De Schrijver and Prinsloo (2000) introduced the concept of 'simultaneous feedback' in lexicography: 'In a nutshell, this new methodology entails the

release of several small-scale parallel dictionaries which triggers feedback that is channeled back to the compilation process of the main dictionaries' (2000: 4). The concept comprises a procedure in which informal and formal criticism on interim versions of a dictionary is collected and used for compiling ever better dictionaries. The method incorporates the target users' needs at an early stage of the compilation of a dictionary.

In dialect lexicography, a variant of 'simultaneous feedback' is sometimes used in order to expand the word material, apart from gaining a better insight in the target users' needs. Many dialect dictionaries present interim dictionary articles to selected respondents for correction or elaboration. Sometimes these articles are published in periodicals on a regular basis, with a request for comments. In the case of the Supplement to the Dictionary of the Zeeland Dialects (2003), for instance, interim versions of the whole dictionary were discussed, word by word, by three groups of respondents. The method proved to be highly effective: apart from many corrections, many new words, meanings, collocations, geographical locations, example sentences, etc. came to light.

1.4.4.6 The structure of the questionnaire

For lexicographic purposes, a thematic arrangement of the questionnaire is advisable because it stimulates the introspection and the memory of the respondent. A questionnaire also has to take the 'filling-in' psychology of the respondent into account. A questionnaire should not be too long; it should contain motivating questions (e.g. easy ones to start with) and should reassure the respondent by pointing out that questions can be skipped if too difficult. Lexemes that easily come to mind should be asked first — these lexemes are not necessarily the ones having the most general meaning. The word for 'very greedy', for instance, is more 'salient' than the word for 'greedy'; hence, the question about the word for 'very greedy' should precede the question about 'greedy', lest the salient words should be given when questions about words with more neutral meanings are asked.

In general, the researcher has to have empathic skills in order to foresee the reactions of the respondents. Filling in a questionnaire should be an agreeable task — although it demands an effort from the respondent's side — and the respondent should have the feeling that the researcher understands the difficulties he may encounter.

1.4.4.7 Fieldwork in practice

The different phases of the fieldwork for the open word classes as — ideally — conducted by the WVD can be summarized as follows:

- (a) Establishment of an inventory of concepts based on
 - the classification of Hallig-Von Wartburg/Frissen/WALD/Van Keymeulen;

- thematic re-arrangement of older dialect dictionaries;
 - systematic inventories of the Dutch standard language (e.g. Brouwers 1988); and
 - older questionnaires.
- (b) Definition of a section in the systematic classification.
 - (c) Preparation of the respondent groups by sending a specimen questionnaire.
 - (d) Processing the answers by the members of the respondent groups in preparation of the group interviews.
 - (e) Interviews with the respondent groups, aiming at
 - enrichment of the inventory of concepts;
 - detection of the structuring principles/structure itself of the semantic field; and
 - collection of words and meanings.
 - (f) Processing the results of the group interviews.
 - (g) Establishment of the definitive questionnaire.
 - (h) Sending out the questionnaire to all the informants.
 - (i) Processing the answers of the investigation by correspondence.

1.5 How much?

The question 'How much?', i.e. the question about the saturation of a collection, is difficult to answer. There is, however, always a point where the continuation of fieldwork (more questions, more respondents, and more localities) is felt to be unproportional to the emergence of new data.

In the case of the WVD, the answer to 'How much?' is related to the answer to 'Where?'. The minimal requirement is met when the WVD has one good piece of lexical information for every lexical area (the lexical aim); the maximal objective is met when there is a good piece of lexical information for every locality, so that accurate lexical maps can be drawn (the geographical aim). In practice, the WVD is totally dependent on the cooperation of volunteers; the geographical aim is hardly ever completely met.

2. Presentation

The presentation of the word material in a dictionary, both macrostructurally and microstructurally, is of course dependent on an analysis of the users' needs. A dictionary should maximize the different usages that are possible in a userfriendly way.

2.1 Presentation of the macrostructure

Thematic versus alphabetical

The choice between a thematic or an alphabetical arrangement of the macrostructure of a dictionary is determined by its purpose. An alphabetical arrangement is of course evident if the dictionary is meant to be merely a tool for looking up the meaning/orthography, etc. of a word. If a dictionary is also meant to be a description of the culture of the language community, a thematic arrangement may be considered. In the latter case, the row of dictionary articles with regard to a certain subject matter can be seen as an inventory of the mental constructs (the meanings) accompanying part of the culture under investigation.

Whatever choice one makes, it is possible to combine the advantages of both arrangements by introducing an alphabetical index to a thematic dictionary or *vice versa* (which is rarely done). It speaks for itself that thematic fieldwork does not prevent a lexicographer from making an alphabetical dictionary. An additional advantage of collecting the lexemes and describing their meanings thematically, is that the lexicographic definitions can be geared to one another, even if they are arranged alphabetically afterwards.

Semasiological versus onomasiological

A semasiologically oriented dictionary tries to describe (among other things) the meaning(s) of lexemes, arranged alphabetically or thematically. The point of departure of an onomasiological dictionary such as the WVD, however, is neither a lexeme, nor a meaning, but the description of a concept, which is defined by a word (or paraphrase) in standard Dutch. Under the so-called standard Dutch 'title' of the dictionary article, all the different dialect words which can be used to refer to the concept are arranged (for an example see Kruijssen and Van Keymeulen 1997: 225-227). It has to be stressed that the description of the concept is not the meaning of all the words listed under it. A word such as *bird*, for instance, may be listed under the concept 'large bird of prey' for those dialects which lack a more specific word.

Classification problems

The assignment of a lexeme to a specific class in a thematic/onomasiological arrangement is not always easy. In the case of the WVD, the principles of functionality and frequency of a concept in the every-day life of the average dialect speaker are used to assign a concept to a particular class (e.g. a tomato is a plant, but its function is to be eaten as a vegetable and it is more frequently thought of as being eaten than grown, hence 'tomato' is in the first place assigned to the class 'food'). In some cases, the same concept is placed in different classes, especially when there are different lexemes corresponding to these

classes. Some Flemish dialects, for instance, use different words for a potato when grown in the garden (*erpel*) or when cooked and eaten (*patat*).

2.2 Presentation of the microstructure

The inclusion of microstructural elements and their presentation is dependent on the overall purpose of the dictionary, based on an analysis of the users' needs. We will only briefly dwell on pronunciation and the written form of the headword.

The extent to which spelling systems for normally unwritten languages exist (and are used in society), vary considerably. If the spelling is not codified, the lexicographer has to be aware that the spelling used for the headwords in his dictionary may be looked upon as 'official' by the users, because it meets a societal need. Hence, the problem of the spelling of the headwords should be looked into very carefully. In the case of the dictionary of the Flemish dialects, the dialectal headwords are 'Dutchified'. Since the phonology of the dialect group under investigation is closely related to the phonology of standard Dutch, and since the Dutch spelling is highly phonological, it is possible to replace the dialect phonemes by their Dutch counterparts, thus 'framing' the dialect words as if they existed in standard Dutch. In fact, many respondents use this method spontaneously, when noting down the dialect words in the questionnaires. In doing so, they bear proof to the fact that Dutchification of the dialectal headwords was a good option. A WVD Dutchified headword is to be considered as a kind of 'reference form' under which a variety of different local dialect pronunciations are subsumed.

3. Conclusions

In the absence of a corpus of written texts, dictionaries of unwritten languages or language varieties are based on lexicographic testimonies of good dialect speakers, collected orally or by correspondence. It is clear that some elements of a possible macro- or microstructure are less conscious in the mind of the speakers than others; most lexical elements of the open word classes can luckily be evoked by introspection and are therefore 'questionable'. The bulk of the lexical data of the major regional dialect dictionaries in the Dutch language area is obtained by questionnaires, especially when they combine a dictionary and a word atlas. Only smaller scale dictionaries can exclusively be based on oral interviews. In many cases word material derived from other sources is only marginal. To my knowledge, no dictionary has been based solely on a corpus of transcriptions of 'free speech'.

Analysing targeted metalinguistical discourse (a series of 'metalinguistical conversations' with groups of trained respondents) in which the linguistic competence of respondents is explored, is in my view a way to gain lexicographic data that are, although not derived from 'spontaneous' speech, adequate for

dictionary making. In such conversations one may proceed from an encyclopedic analysis of reality itself to the way the world is lexically categorized and further to the collection of lexical data. I think that a thematic lexicographic investigation is very productive, because the researcher can focus on the content of a theme and the completeness of a dictionary with regard to specific subjects can also be assessed. Questionnaires may complete, expand and verify the data gained in interviews. A dictionary based on a corpus of texts is sometimes thought to be more scientific than a dictionary based on the introspection of the lexicographer or respondents. Fieldwork can of course be carried out scientifically as well. In the last analysis, it is the user who is the best judge of the quality of a dictionary and his judgement should be systematically enlisted at a very early stage. The method of 'simultaneous feedback' is perhaps even more desirable for fieldwork dictionaries than for corpus-based dictionaries, to overcome the danger of the subjectivity or the incompleteness of collection methods based on linguistic introspection.

In the case of the Dictionary of the Flemish Dialects and its two cognate dictionaries, the thematic approach, both in collecting and presenting the word material, undoubtedly rescued the project financially. Since it was possible to publish a thematic fascicle every year, the editorial board managed to maintain the project financially. Nobody but the lexicographer understands why it takes such a long time to compile a dictionary. The sponsors have to be kept happy.

References

- Ammon U, N. Dittmar and K. Mattheier (Eds.).** 1988. *Sociolinguistics. An International Handbook of the Science of Language and Society./Soziolinguistik. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft.* Berlin/New York: De Gruyter.
- Atteslander, P.** 1988. Befragung. Ammon U, N. Dittmar and K. Mattheier (Eds.) 1988: 940-951.
- Besch, W., U. Knoop, W. Putschke and H.E. Wiegand (Eds.).** 1982/1983. *Dialektologie: ein Handbuch zur Deutschen und Allgemeinen Dialektforschung.* Handbücher zur Sprach- und Kommunikationswissenschaft 1,1 and 1,2. Berlin/New York: De Gruyter.
- Brouwers, L.** 1988. *Het juiste woord.* Antwerpen: Standaard Uitgeverij.
- Cicourel, A.** 1988. Elicitation as a Problem of Discourse. Ammon U, N. Dittmar and K. Mattheier (Eds.). 1988: 903-910.
- Denzin N. and Y. Lincoln (Eds.).** 1994. *Handbook of Qualitative Research.* Thousand Oaks/London/New Delhi: Sage Publications.
- De Schrijver, Gilles-Maurice and D.J. Prinsloo.** 2000. The Concept of "Simultaneous Feedback": Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.
- Fontana, A. and J. Frey.** 1994. Interviewing. The Art of Science. Denzin, N. and Y. Lincoln (Eds.). 1994: 361-376.
- Frissen, C.** 1981. *Concept Inventory for WBD/WLD. Result of the Project "Vorbereidend onderzoek in verband met de algemene woordenschat" (Preparatory Investigation Concerning the General Vocabulary) 1978-1981.* Unpublished Typescript. Nijmegen: Catholic University.
- Hallig, R. and W. von Wartburg.** 1952. *Begriffssystem als Grundlage für die Lexikographie.* Abhand-

lungen der deutschen Akademie der Wissenschaften zu Berlin. Klasse der Sprachen, Literatur und Kunst nr. 4. Berlin: Akademie Verlag.

König, W. 1982/1983. Probleme der Repräsentativität in der Dialektologie. Besch, W., U. Knoop, W. Putschke and H.E. Wiegand (Eds.). 1982/1983: 463-485.

Kruijssen, J. en J. van Keymeulen. 1997. The Southern Dutch Dialect Dictionaries. *Lexikos* 7: 207-228.

Merton, R., M. Fiske and P. Kendall. 1956. *The Focused Interview*. Glencoe, IL: Free Press.

Van Keymeulen, J. 1992. *De algemene woordenschat in de grote dialectwoordenboeken (WBD, WLD, WVD). Een methodologische reflectie*. Ph.D. Dissertation. Ghent: Ghent University.

Van Keymeulen, J. 2003a. Dialectwoorden verzamelen. Een praktische handleiding. *Handelingen van de Koninklijke Commissie voor Toponymie en Dialectologie* 75: (forthcoming).

Van Keymeulen, J. 2003b. De introductie van het Nederlands in Vlaanderen: taalstrijd van de middenklasse. Botha, Willem (Ed.). *'n Man wat beur. Huldigingsbundel vir Dirk van Schalkwyk*: 332-339. Stellenbosch: Bureau of the WAT.

WALD = Schaars, A. 1984-. *Woordenboek van de Achterhoekse en Liemerse Dialecten*. Doetinchem: Staring Instituut.

WBD = Weijnen, A. et al. 1967-. *Woordenboek van de Brabantse Dialecten*. Assen: Van Gorcum.

Weijnen, A. en J. van Bakel. 1967. Voorlopige Inleiding op het Woordenboek van de Brabantse Dialecten. WBD. 1967-.

WLD = Weijnen, A. et al. 1983-. *Woordenboek van de Limburgse Dialecten*. Assen: Van Gorcum.

WVD = Devos, M. et al. 1979-. *Woordenboek van de Vlaamse Dialecten*. Tongeren: Michiels.