

> **State of the Art of the Lexicography of European Lesser Used or Non-State Languages**

ANNE TJERK POPKEMA

‘The people who chronicle the life of our language (...) are called lexicographers’ (Martin Hardee, blogger in Cyberspace, 2006)

o **Introductory remarks<sup>1</sup>**

Language codification and language elaboration (‘Ausbau’) are key ingredients for raising a lesser used language to a level that is adequate for modern use.<sup>2</sup> In dictionaries (as well as in grammars) a language’s written standard may be laid down, ‘codified’.<sup>3</sup> At the same time dictionaries make clear what lexical gaps remain or arise in a language. The filling of such gaps – part of language elaboration – will only gain wide acceptance when, in turn, it is codified in a dictionary itself. Thus, both prime categories of language development – codification and elaboration – are hats worn by the same head: the lexicographer’s.

Bo Svensén begins the opening chapter of his recent handbook on lexicography by stating that ‘dictionaries are a cultural phenomenon. It is a commonplace to say that a dictionary is a product of the culture in which it has come into being; it is less so to say that it plays an important part in the development of that culture.’<sup>4</sup> In the case of lesser used languages, language development may lead to (increased) use in domains that were formerly out of reach because of the dominance – for any number of reasons – of another language. In such instances, language development equals language emancipation. An emancipating language takes on new functions, enters new domains of society and is therefore in need of new terminology. As a result, the language needs new or revised dictionaries that in their turn strengthen the ongoing language emancipation – a virtuous circle with the lexicographer at its very heart. Such a circle may turn vicious just as easily, when emancipatory efforts are unsuccessful

---

1 I am grateful to Anne Dykstra, Willem Visser (both: Fryske Akademy Leeuwarden, the Netherlands) and Alastair Walker (Nordfriesische Wörterbuchstelle, Christian-Albrechts-Universität Kiel, Germany) for their comments on an earlier version of this paper.

2 Cf. Haugen (1966:931)

3 Cf. Inoue (2006): ‘One concrete strategy of minority language revitalization is formal codification through practices of literacy, orthography, dictionary, grammar, or census.’

4 Svensén (2009:1)

or even absent, and domains and functions are firmly held or taken over by the dominant language. Yet in this case, too, the lexicographer finds himself at the heart of the circle, He documents as much as he can of the language in decline, objectively executing his scientific, descriptive task (although he may well have silent hopes of stopping or slowing down the downwards spiral).

In both capacities, the lexicographer is a language emancipator – whether he wants to or not. For even the fiercest denier of language-ideological influence on his lexicographic activities cannot prevent his strictly objective language description – his dictionary – becoming a tool of language emancipation. Even more so: the more objective, the more ‘scholarly’ a dictionary of a language is, the more up to date and elaborate it may become, and the more it may be deployed by language movements in their efforts to elevate a language’s social status: ‘It is certainly a real language – look at this enormous dictionary!’ It’s an example of the classic paradox of influencing an object of research by studying it objectively.

The capacity of lexicography as a prime emancipatory tool for lesser used languages makes the lexicographer a key figure in the play of language life – and, alas, at times a close spectator of the tragedy of language death. For undeniably, a lexicographer of an endangered language runs the increased risk of finding himself at the language’s death bed, meticulously documenting its final gasps of air. Under such circumstances, the lexicographer’s task is of equal importance as it is for revitalizing purposes: a language contains centuries, possibly millennia of cultural and ecological information that may be forever lost if it is not documented properly. In 2003, UNESCO language experts stated that ‘a language that can no longer be maintained, perpetuated, or revitalized still merits the most complete documentation possible. This is because each language embodies unique cultural and ecological knowledge in it. It is also because languages are diverse. Documentation of such a language is important for several reasons: 1) it enriches the human intellectual property, 2) it presents a cultural perspective that may be new to our current knowledge, and 3) the process of documentation often helps the language resource person to re-activate the linguistic and cultural knowledge.’<sup>5</sup>

Witnessing language extinction certainly is a gloomy perspective, yet for

---

5 UNESCO (2003), par. 3.5, at <http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf>.

many lexicographers of endangered languages across the globe it is not an entirely unlikely professional destiny. Such lexicographers often operate in relative isolation – the poorer the patient, the less doctors he’s likely to see. And, like any friendly small town physician, the lexicographer is often not left unmoved by the decline. For although lexicography of lesser used languages is a full-grown scientific discipline and most lexicographers are well capable of objectively documenting the language concerned – still, in my experience many lexicographers of a lesser used language tend to not to be fully objective towards the language. Not uncommonly, they are either native speakers – and who wishes one’s mother tongue to become extinct? – or they have come to appreciate the language they document – and who remains unmoved by a dear friend’s good or misfortune?

In my opinion, however, a certain degree of subjectivity with respect to a language’s vitality does not imply an unscientific attitude by definition. It may well often be such personal involvement that allows the lexicographer to get close enough to the language speakers to be able to document their language in its entirety. An objective, hard-core scientific attitude may be perceived as ‘cold-hearted’ by the native speakers, and might lead to less than full openness, which in turn may lead to less than optimal research results – a poorer dictionary.

It would, however, appear that relative professional isolation and personal involvement are more of a rule than an exception in the lexicography of lesser used languages, making exchanging information and experience with fellow-lexicographers useful, even vital at times. In the light of such considerations and, indeed, responsibilities, it is fortunate that the organizing committee of the Euralex International Congress chose to make the lexicography of lesser used or non-state languages the central theme of their 14th gathering.

One important feature of the conference is a survey of the state of the art of the lexicography of such languages. To this end, the organizing committee sent out questionnaires to dozens of institutions and individuals involved in the lexicography of lesser used or non-state languages. In this paper, some of the results of this Euralex 2010 Survey of the Lexicography of European Lesser Used or Non-State Languages (henceforth: Survey) are presented.

In the first section of this paper, I will discuss several methodological aspects of the questionnaire and the Survey. The second section is the core of this paper, providing facts and figures, including indispensable

sociolinguistic information on the languages in the Survey and summarized results of various lexicographical aspects. In the third section some implicational statements are made. Since offering and analyzing all the data the Survey has culled in the field of lexicographical practice (and there is quite a considerable amount – it will hopefully be made generally available via the Euralex website) was simply impossible because of the limited space for this paper, I chose to concentrate on some of the results I found particularly interesting in the third section. The fourth section consists of a few final remarks.



*Map 1: Geographical distribution of all lesser used or non-state language areas of Europe*

## **1 The questionnaire<sup>6</sup>**

### *European lesser used or non-state languages*

In order to give an impression of the scale of European lesser used or non-state languages, perhaps I might quote some figures.<sup>7</sup> Except for Iceland, every European state has at least one linguistic minority (cf. Map

---

6 On language surveys – including those on minority languages – and their pitfalls, cf. De Vries (2006).

7 All numbers of speakers and percentages mentioned in the following sections are approximate, without explicitly stating so each time.

1). The total number of lesser used or non-state European languages is approximately 60, representing 55 million European citizens.<sup>8</sup> Keeping these numbers in mind as well as the above-mentioned importance of lexicography for language development and language emancipation, it seems rather odd that up till now there has been no comparative overview of the lexicographical situation of lesser used or non-state languages in any lexicographical or (socio)linguistic handbook on European languages.<sup>9</sup> There are numerous overviews of the lexicographical state of affairs of individual lesser used or non-state languages as well as overviews of the lexicography of such languages within a single state's borders,<sup>10</sup> yet an overview transcending the state level has, to the best of my knowledge, never been attempted.

*Terminology I: 'lesser used or non-state languages'*

A particular minefield is the nomenclature in the semantic range of

---

8 Cf. Facts and Figures on the website of Mercator European Research Centre on Multilingualism and Language Learning (henceforth: Mercator), at <http://www.mercator-research.eu/minority-languages/facts-figures>.

9 Such (socio)linguistic handbooks include: Hinderling/Eichinger (1996); Janich/Greule (2002) (which includes paragraphs on the lexicography of individual languages, in which the most important dictionaries are mentioned); Åkermark e.a. (2006). The 16<sup>th</sup> online edition of Ethnologue Languages of the World, at <http://www.ethnologue.com/web.asp>) also provides a lot of sociolinguistic information on virtually all languages of the world. Sociolinguistic data on dozens of European minority languages are available in the Introductions to the Mercator Regional Dossiers. Among these are many on languages represented in this Survey (viz. Asturian (in Spain), Basque (both in France and in Spain), Catalan (both in France and in Spain), West Frisian, Galician (in Spain), Scottish Gaelic, Latgalian, North Frisian, Sami (in Sweden), Sorbian (including Lower Sorbian), Võro and Welsh). All of Mercator's Regional Dossiers are digitally accessible at <http://www.mercator-research.eu/research-projects/regional-dossiers>).

10 The outstanding lexicographic handbook Hausmann e.a. (1989-1991) offers excellent overviews for some of the languages present in the Survey (cf. Table 1), viz. Galician (article nr. 181a), Catalan (184), Romansh (190), the Frisian languages (202), the Sorbian languages (210), Basque (226) and the Sami languages (228a). The lexicography of Romance languages is well covered in the monumental LRL (the Romance languages in the Survey that are treated in the LRL are: Friulian (LRL III, art. 217), Romansh (III 233b), Catalan (V, 2 358b), Asturian (VI, 1 408, pp. 688-689) and Galician (VI, 2 417)). Also, the equally monumental ELL offers overviews on the lexicography of many languages in the Survey (Welsh, by Hawke (2006), Frisian, by Bremmer (2006), Nynorsk, by Kulbrandstad/Veka (2006, par. 'Norwegian'), Galician, Asturian, Catalan and Basque (combined in Saurí Colomer (2006)). ELL also contains an overview of lexicographical topics and issues by Hanks (2006) including a short paragraph on 'Dictionaries of Rare and Endangered Languages', providing merely a few general remarks on tools and aims of such dictionaries.

‘language and dialects’ – the linguist’s approach to what constitutes a dialect may differ entirely from that of a sociolinguist or a language policy maker.<sup>11</sup> In the Survey as well as in this paper, the term ‘lesser used or non-state languages’ is used in a sense similar to the term ‘regional or minority languages’ as defined in the *European Charter for Regional or Minority Languages* (henceforth: Charter).<sup>12</sup> Such lesser used or non-state languages include: unique languages in one state (e.g. West Frisian in the Netherlands); unique languages spread over more than one state (e.g. Basque in Spain and France); transfrontier languages that are both minority and majority languages, depending on the state (e.g. Sweden Finnish); non-territorial languages (e.g. Romani or Yiddish); official languages that are lesser used on the whole or part of the state’s territory (e.g. Romansh in Switzerland).<sup>13</sup>

#### *Terminology II: ‘dictionary’ vs ‘wordlist’*

Special attention should also be paid to the problem of defining the concepts ‘dictionary’ and ‘wordlist’, which are both key concepts in the questionnaire. The grey area between the two types is both vast and treacherous. Not every informant will distinguish equally sharply or consistently between the two, which is partly due to the fact that ‘dictionary’ does not only denote a specific lexicographical type, but also

---

11 Cf. Haugen (1966), esp. pp. 926-927.

12 See: Charter, Part I art. 1a and art. 3 sub 1, at <http://conventions.coe.int/treaty/en/Treaties/Html/148.htm>. The term ‘regional or minority languages’ is avoided here since the organizing committee deemed ‘lesser used or non-state languages’ to be less politically charged. Also, seemingly, there is somewhat of a paradox in the Charter definition, as it aims at protecting languages that are ‘different from the official language(s) of that State’. However, it does explicitly protect some lesser used languages that are in fact official state languages. This is the case, for example, with Romansh and Italian in Switzerland and with Swedish in Finland, which all are official state languages, yet nevertheless have been brought up for protection under the Charter by the Swiss resp. Finnish government. Cf., however, footnote 13.

13 For the latter category, cf. paragraph 51 of the *Explanatory Report* on the Charter (at <http://conventions.coe.int/treaty/en/Reports/Html/148.htm>): ‘The wording of Article 3 takes account of the position in certain member states whereby a national language which has the status of an official language of the state, either on the whole or on part of its territory, may in other respects be in a comparable situation to regional or minority languages as defined in Article 1, paragraph a, because it is used by a group numerically smaller than the population using the other official language(s).’ For comments on terminology, also cf. the *Explanatory Report*, art. 18. I am grateful to Auke van der Goot, employee of the Dutch Ministry of Interior, for pointing this out to me.

is an umbrella term for several lexicographical types.<sup>14</sup> The latter usage explains the tendency among many publishers (and authors!) to name any lexicographical product ‘dictionary’, which does not, however, help untangle matters. It is in fact somewhat of a paradox that the number of wordlists (or genres related in magnitude and scope like glossaries, lexica, vocabularies) by far exceeds that of actual dictionaries, yet the title words ‘wordlist/glossary/vocabulary/lexicon’ are by no means as numerous as the title word ‘dictionary’, which is obviously all too often used in its umbrella-term capacity.<sup>15</sup> Thus it takes quite a theoretical lexicographer to draw a sharp and consistent line between the lexicographical types, and there is no harm in acknowledging that not all informants have been equally successful in doing so.

#### *Aim and set-up*

The main aim of the Survey is to ascertain the current state of affairs in the lexicography of European lesser used or non-state languages and their social and linguistic situation.<sup>16</sup> To this end a questionnaire was compiled by the organizing committee (see Appendix).<sup>17</sup>

The questionnaire consists of three main parts. In the first part, contact data for the individual or organization responsible for filling out the questionnaire are gathered. The second part consists of questions on the sociolinguistic position of the language concerned (e.g. questions on the numbers of speaker (2.3), regions/states in which the language is spoken (2.2), related state languages (2.4), the level of recognition by the national government (2.5), the existence of an official spelling (2.11) and grammar (2.13)). The third part is divisible into two subparts. The first subpart (3.1-

---

14 Cf. Hartmann/James (1998), s.v. *dictionary*. For anyone not quite familiar with the distinction between dictionaries proper vs wordlists and the like, I would recommend reading Simpson (1993, esp. pp. 124-125), who provides a very accessible introduction for laymen, taking lexicography of Aboriginal languages as a starting point.

15 A quick search on international bookselling site Amazon.com rendered 180,327 titles containing the word ‘dictionary’ and a mere total of 78,290 titles containing either ‘wordlist’ (385) or ‘vocabulary’ (32,232) or ‘glossary’ (13,522) or ‘lexicon’ (7,185). A Google quick search for the same words (which renders a completely different range of hits, since not only book titles are found), still shows ‘dictionary’ as the more popular term over the other ones combined (139,000,000 vs 125,430,000 hits).

16 The Survey was set up in close collaboration with Mercator, which is affiliated to the Fryske Akademy in Leeuwarden, The Netherlands.

17 The questionnaire in the Appendix represents a slightly adapted version of the original questionnaire. The adaptations merely concern matters of lay-out; the phrasing of the questions has not been altered. The questionnaire can also be downloaded via the Euralex website (<http://www.euralex2010.eu/>).

3.23) contains questions on aspects of the lexicographical situation of the language (e.g. the existence of monolingual or bilingual dictionaries or wordlists, the way in which they were published, numbers of sales, the existence of lexicographic tools in education). The second subpart (3.24-3.35) is aimed at gathering information on the lexicographic infrastructure (e.g. questions on subsidies for compiling dictionaries (3.26) or on support for setting up and maintaining a lexicographical infrastructure (3.27), the embedding of lexicography in other linguistic research (3.30), the way in which primary sources of the language are organized and processed (3.32-3.33)).

### *Response*

The organizing committee put a lot of effort into contacting as many institutions or individuals involved in the lexicography of lesser used or non-state languages as possible. To this end the Mercator Database of Experts proved especially useful.<sup>18</sup> In the event of languages for which experts were lacking in the Mercator Database of Experts, the committee tried complementing the list of informants by making use of the Mercator Regional Dossiers,<sup>19</sup> which often contain useful contact information.

These efforts led to fully satisfactory results. As mentioned above, European lesser used or non-state languages are extant in all European countries, except for Iceland. Their total number is approximately 60, representing a total of 55 million European citizens. The committee received a total of 26 completed questionnaires on European languages (as well as 4 on non-European languages, see below). The total amount of European languages present in the Survey is 22, covering 19 Mio speakers. Hence, the Survey covers well over 1/3rd of all European lesser used or non-state languages and of the total number of their speakers.

The response shows that, although questionnaires were sent out to well over a dozen of lesser used languages that are a dominant state language elsewhere, only one response came in, namely from Sweden Finnish. By comparison, the response ratio for languages that are not dominant elsewhere is approximately 2:1, with two responses for every lack of response.

---

18 At <http://www.mercator-research.eu/minority-languages/database%20of%20experts>. Tjallien Kalsbeek (Mercator) was especially helpful in compiling the extensive list of informants, and I am grateful to Liesje Haanstra (Fryske Akademy) for collecting and forwarding the completed questionnaires.

19 Cf. footnote 9.

Although the Survey primarily aimed at gathering data on European lesser used or non-state languages, several informants of non-European lesser used languages were also approached. Response was received for Nivkh (Russian peninsula of Sakhalin, from two informants, independent of each other), for a group of minor East Iranian languages and for South Efate (Polynesian isle of Vanuatu).

The committee received a total of 30 completed questionnaires. Unavoidably, in order to achieve a survey of European lesser used or non-state languages, it proved necessary to make a selection of the questionnaires. Firstly, since a survey of European languages is concerned here, questionnaires on non-European languages were not taken into account (these are, as already mentioned: Nivkh, a group of minor East Iranian languages and South Efate).<sup>20</sup> Also, for some European languages duplicates were handed in, namely on Galician and Catalan. The data from such parallel questionnaires were combined into one new, 'optimized' version. Furthermore, in two instances, questionnaires were filled out for varieties of a language that was already present in the response (viz. Helgolandic (North Frisian) and Algherese (Catalan)). The data from these questionnaires were incorporated into the questionnaire of the language of which they are varieties. Sweden Finnish was also left out. As mentioned earlier, it is the only language which is a dominant state language elsewhere for which a questionnaire was filled out. For its lexicographical infrastructure and output Sweden Finnish may rely on the lexicographical infrastructure of its 'homeland' Finland. This renders it less useful for the purpose of comparing languages that do not have such an infrastructure to rely on.<sup>21</sup>

As a result of the selection criteria, in total 22 questionnaires were deemed suitable for processing in the Survey (see Map 2 for their geographical distribution).

#### *Processing the data*

The questionnaire contained over a hundred questions (cf. Appendix). As it turned out, some questions were hard to answer, either because the answering options were insufficient, or because the question itself did

---

<sup>20</sup> The data these questionnaires contain will not be neglected. In due time Euralex hopes to be able to present the questionnaires on their website.

<sup>21</sup> Actually, Catalan may also count as a language that is a dominant state language elsewhere: it is the only official state language of Andorra. However, the lexicographical epicentre of Catalan is located in Spanish Catalonia. For this reason, Catalan was not excluded.

not (fully) apply.<sup>22</sup> Also, several questions were skipped by a large portion of the informants since they did not have relevant data at their immediate disposal (this applies particularly to several questions in the third section, on URL's and on printing and sales numbers) or since answering them would simply take up too many resources (e.g. extensive bibliographical information). As a result, there were several questions that (might have) rendered unreliable data. They were filtered out for that reason.

As it turned out, some questions were answered incompletely, by a small number of informants or in a heterogenic way. Such questions include the ones on the manner in which dictionaries and wordlists are published (3.25), on their goal or background (3.20) and on (the lack of) non-governmental subsidies (3.26). Making general statements on the basis of such results would not render reliable information. Hence, such questions were left out.

It was rather unfortunate that in some cases the questionnaire did not offer sufficient opportunity for specifying information for individual dictionaries and wordlists. For example, following a list of bilingual dictionaries, the informant was asked whether they are online, which may differ per dictionary. Yet, by offering merely one 'yes/no' choice box (3.6), the informant was unfortunately not provided the opportunity to specify per dictionary. The overall decision was made to regard such questions as answered with 'yes' if the answer applied to any of the listed publications, and with 'no' if nothing was indicated at all.

There is also the above-mentioned matter of terminology: not every informant will use or define key-concepts like 'wordlist' or 'dictionary' in the same way. Quite often, 'dictionary' was used as an umbrella term for any lexicographical type. Of course, questions containing these concepts can hardly be filtered out in a survey on lexicography. Here a rather pragmatic approach was followed: where a questionnaire was deemed to be falsely indicating a wordlist as a dictionary or vice versa (which occurred much less frequently), the answer was corrected silently. It should be noted, however, that the questionnaires were not systematically checked with respect to the dictionaries and wordlists mentioned.

---

22 For example: the question on the tendency of distancing (question2.17) is often not relevant for languages that are genetically distant to the dominant language. In the case of (Slavonic) Lower Sorbian in Germany, for instance, this is the case, and therefore distancing is not a relevant issue here (Gunter Spiess (formerly) of the Sorbisches Institut in Cottbus, Germany, in personal correspondence with the author).

There seems to have been some confusion as to whether the questionnaire should be filled out for lexicographical projects the informant is personally involved in or for all projects of the language of the informant's expertise. The latter was the aim of the committee and, although this was not explicitly indicated anywhere, fortunately most informants seem to have filled out the questionnaire accordingly.

Information on protection under the Charter as well as on the level of endangerment according to the UNESCO Atlas of Languages in Danger (henceforth: UNESCO Atlas) was added by myself at a later stage.<sup>23</sup>

## 2 Facts and Figures

The main part of the questionnaires consisted of questions on lexicographic infrastructure and lexicographical output of the language. In order to evaluate such topics, the languages need to be placed in some sociolinguistic perspective. Numbers of speakers of the languages, protection under the Charter, level of endangerment, countries in which they are spoken: all of these may influence a language's social status, which in turn may have an effect on lexicographical practice and possibilities. Therefore, the following pages contain several sets of thematically linked results of the Survey. The initial two sets consist of sociolinguistic data. The first set (A) is on the states in which the languages are spoken, the state borders they cross and their numbers of speakers. The second set (B) is concerned with governmental recognition (including protection by the Charter) and the level of endangerment. After this, two sets of facts and figures concerned with lexicography follow: the first one (C) focuses on lexicographic infrastructure and the latter (D) on actual lexicographical output. Section (E) contains two tables showing, respectively, the level of diversity of lexicographical output per language, and the level of use of modern technology in lexicographic practice.

The information offered is restricted to mere facts. What the data might tell us, what they mean or why they are what they are, is the content of paragraph 3 ('Implicational Statements').<sup>24</sup>

---

<sup>23</sup> At <http://www.unesco.org/culture/ich/index.php?pg=00206>.

<sup>24</sup> I stress that the data that are listed on the next few pages represent only part of what the combined questionnaires have to offer: as a result of the limited space only a selection of the results can be presented here.

<b>Language (state(s))</b>	<b>No. of speakers</b>
(A) Asturian (Spain/Portugal)	350,000-500,000
(B) <i>Basque (Spain/France)</i>	1 Mio
(C) <i>Catalan (Spain/France/Italy/Andorra)</i>	7-9 Mio
(F) Friulian (Italy)	600,000
(FN) <i>North Frisian (Germany)</i>	8,000
(FS) <i>Sater Frisian (Germany)</i>	2,000
(FW) <i>West Frisian (Netherlands)</i>	450,000
(G) <i>Galician (Spain/Portugal)</i>	2 Mio
(J) Jèrriais (UK)	2,600
(L) Latgalian (Latvia)	150,000-200,000
(LG) <i>Low German (Germany)<sup>1</sup></i>	2.6 Mio
(LSa) <i>Low Saxon (Netherlands)<sup>2</sup></i>	2.15 Mio
(LSo) <i>Lower Sorbian (Germany)</i>	7,000
(NN) <i>Nynorsk (Norway)</i>	500,000
(R) <i>Romansh (Switzerland)</i>	60,000
(SG) <i>Scottish Gaelic (UK)</i>	60,000
(SI) <i>Inari Sami (Finland)</i>	300
(SK) <i>Kildin Sami (Russia)</i>	300-700
(SN) <i>North Sami (Norway/Finland/Sweden)</i>	30,000
(SS) <i>Skolt Sami (Finland/Russia)</i>	300
(V) Võro (Estonia)	50,000-70,000
(W) <i>Welsh (UK)</i>	600,000

Table 1: languages represented in the Survey (cf. Map 2)  
(in italics: protected under the Charter)

**A1.** Lesser used or non-state languages are extant in all European countries, except for Iceland (cf. Map 1). Their total number is approximately 60, representing a total of 55 Mio European citizens. The total number of languages in the Survey is 22, covering 19 Mio speakers in 15 different countries.

**A2.** There are 7 transfrontier languages in the Survey: Asturian (Spain/Portugal), Basque (Spain/France), Catalan (Spain/France/Italy/Andorra), Galician (Spain/Portugal), Low German/Low Saxon (Germany/Netherlands),<sup>25</sup> North Sami (Norway/Finland/Sweden) and Skolt Sami (Finland/Russia).



Map 2: geographical distribution of European languages present in the Survey  
(for the abbreviations, cf. Table 1)

25 In the Survey, Low Saxon and Low German are both considered to be transfrontier languages. This does, however, call for some elaboration. Due to increasing influence in the course of the 20th century of Dutch (west of the state border) and High German (east of the border) (cf. Niebaum (2008a) and (2008b, esp. 437-438), a formerly relatively coherent language area diverged into two sets of dialect groups: a Dutch one, Low Saxon, and a German one, Low German. As a result, the Dutch-German state border does constitute a linguistic border nowadays: it divides the larger coherent language area, stretching from the Veluwe region at the heart of the Netherlands all the way east to the German-Polish state border. Nevertheless, these two larger dialect groups do still undeniably constitute a transfrontier dialect continuum, be it with a sharper linguistic division than some decades ago. To illustrate the taxonomical confusion: The UNESCO Atlas (cf. footnote 23) considers ‘Low Saxon’ (with the alternate name: Low German) to be a language spoken both in the Netherlands and in Germany, as does Ethnologue (cf. footnote 9).

**A3.** Out of the 19 Mio speakers covered by the Survey, almost 60% are Spanish citizens. With 8 Mio speakers (42% of the total), Catalan is by far the largest language in the Survey. Taking Catalan out of the equation would still mean that 32% of the remaining population represented by the Survey is of Spanish nationality.

**A4.** Skolt Sami (300 speakers), Inari Sami (300) and Kildin Sami (500) are the languages in the Survey with the smallest numbers of speakers.<sup>26</sup>

**A5.** In the Survey, the two largest language groups are Germanic and Romance languages: both categories are represented by 6 languages, covering 5.7 Mio (Germanic) and 11 Mio speakers (Romance).<sup>27</sup>

#### *B – Sociolinguistic: Governmental Recognition and Level of Endangerment*

**B1.** All languages in the Survey have in some way been officially recognized by their national governments – the only exception is Võro (Estonia).<sup>28</sup> All languages also receive financial support from the government.<sup>29</sup>

**B2.** Out of the 15 states represented in the Survey by one or more languages, 7 have not ratified the Charter.<sup>30</sup> As a result, Friulian (Italy), Latgalian (Latvia), Kildin Sami (Russia) and Võro (Estonia) are not under its protection. Furthermore, two languages (Jèrriais (UK) and Asturian

---

26 The fourth Sami language in the Survey, North Sami, is spoken by 30,000 people.

27 Germanic: Low German, Low Saxon, West Frisian, Sater Frisian, North Frisian, Nynorsk. Romance: Catalan, Asturian, Galician, Jèrriais, Friulian, Romansh.

28 However, Võro does receive governmental financial aid for lexicographical purposes, which is seemingly a bit of a paradox. On enquiring about this matter, I received the following comments: ‘Võro is the most vivid variant of what is now known as South Estonian. Lately, South Estonian is officially indicated as ‘a special modification of Estonian’, but not as a language. The Estonian government does acknowledge that South Estonia and the West coast islands are culturally special and in need of support. The Ministry of Culture hosts several programs for cultural support of South Estonian, out of which one program is especially meant for Võro. Furthermore, Võro is supported by funds from the Ministry of Education (as part of the program ‘Estonian and National Memory’) for lexicographical work. Also, the Võro Institute is a state institution, which receives state funds from the Ministry of Culture.’ (Mariko FASTER of the Võro Institute in Võro, Estonia, in a personal email to the author, February 2010).

29 In some cases, this may be a local or regional rather than the national government.

30 Viz. Andorra, Estonia, France (signed, not ratified), Italy (signed, not ratified), Latvia, Portugal, Russia (signed, not ratified).

(Spain/Portugal)) are spoken in countries which did ratify the Charter,<sup>31</sup> yet were not brought up for protection under the Charter, bringing the total of languages not protected by the Charter to 6.<sup>32</sup> The combined number of speakers of these 6 languages is 1.2 Mio.

<b>Languages (no. of speakers)</b>	<b>Level of endangerment</b>	<b>Level of endangerment</b>	<b>Languages (no. of speakers)</b>
Asturian (350-400,000)	Definitely endangered	(Not listed)	<i>Catalan</i> (7-9 Mio)
<i>Basque</i> (1 Mio)	Vulnerable	(Not listed)	<i>Galician</i> (2 Mio)
<i>Catalan</i> (7-9 Mio)	(Not listed)	(Not listed)	<i>Low German</i> (2.6 Mio)
Friulian (600,000)	Definitely endangered	(Not listed)	<i>Nynorsk</i> (500,000)
<i>Galician</i> (2 Mio)	(Not listed)	Vulnerable	<i>Basque</i> (1 Mio)
<i>Inari Sami</i> (300)	Severely endangered	Vulnerable	<i>Latgalian</i> (150-200,000)
Jèrriais (2,600)	Severely endangered	Vulnerable	<i>Low Saxon</i> (2.15 Mio)
Kildin Sami (300-700)	Severely endangered	Vulnerable	<i>Welsh</i> (580,000)
<i>Latgalian</i> (150-200,000)	Vulnerable	Vulnerable	<i>West Frisian</i> (450,000)
<i>Low Saxon</i> (2.15 Mio)	Vulnerable	Definitely endangered	Asturian (350-400,000)
<i>Low German</i> (2.6 Mio)	(Not listed)	Definitely endangered	Friulian (600,000)
<i>Lower Sorbian</i> (7,000)	Definitely endangered	Definitely endangered	<i>Lower Sorbian</i> (7,000)
<i>North Frisian</i> (8,000)	Severely endangered	Definitely endangered	<i>North Sami</i> (30,000)
<i>North Sami</i> (30,000)	Definitely endangered	Definitely endangered	<i>Romansh</i> (60,000)
<i>Nynorsk</i> (500,000)	(Not listed)	Definitely endangered	<i>Scottish Gaelic</i> (60,000)
<i>Romansh</i> (60,000)	Definitely endangered	Definitely endangered	<i>Võro</i> (50,000-70,000)
<i>Sater Frisian</i> (2,000)	Severely endangered	Severely endangered	<i>Inari Sami</i> (300)
<i>Scottish Gaelic</i> (60,000)	Definitely endangered	Severely endangered	Jèrriais (2,600)
<i>Skolt Sami</i> (300)	Severely endangered	Severely endangered	Kildin Sami (300-700)
<i>Võro</i> (50,000-70,000)	Definitely endangered	Severely endangered	<i>North Frisian</i> (8,000)
<i>Welsh</i> (580,000)	Vulnerable	Severely endangered	<i>Sater Frisian</i> (2,000)
<i>West Frisian</i> (450,000)	Vulnerable	Severely endangered	<i>Skolt Sami</i> (300)

*Table 2: Level of endangerment of languages in the Survey according to the UNESCO Atlas (in italics: under protection of the Charter)*

31 For Asturian, this only holds true for Spain. Portugal has neither signed nor ratified the Charter.

32 In the case of Jèrriais, the following applies: ‘The situation for Jèrriais is slightly unclear at present. It seems that all technical obstacles and objections have been answered, and Jersey has received no objection on constitutional grounds from the UK to a ratification on Jersey’s behalf. What remains is the political situation in Jersey (...) However, in September 2009 the Minister for Education, Sport and Culture signed an agreement to formalize the remit of L’Office du Jèrriais in terms of the promotion of Jèrriais and of drawing up language plans with States authorities. L’Office du Jèrriais is therefore now tasked with acting as though ratification has been completed.’ (Geraint Jennings of L’Office du Jèrriais, in a personal email to the author, February 2010).

**B3.** Asturian is the only transfrontier language in the Survey that is not protected by the Charter.

**B4.** Whereas all Germanic languages in the Survey are protected by the Charter, only half of the Romance ones are (viz. Romansh, Catalan and Galician. Not protected are: Asturian, Friulian, Jèrriais).

Not included in the questionnaire was the language's level of endangerment according to the UNESCO Atlas.<sup>33</sup> It contains 5 different levels of language endangerment: vulnerable, definitely endangered, severely endangered, critically endangered, extinct. Table 2 shows how the languages in the Survey are categorized according to the UNESCO Atlas.

**B5.** Table 2 shows that 9 languages in the Survey are considered either vulnerable or are not listed,<sup>34</sup> whereas the remaining 13 languages are considered to be endangered. Out of these 13, 6 are severely endangered<sup>35</sup> No languages in the Survey are considered critically endangered.

**B6.** Except for Latgalian, all languages that are not under protection of the Charter are considered either definitely or severely endangered.<sup>36</sup>

#### *C – Lexicographical: Infrastructure*

**C1.** Although all languages in the Survey receive governmental support (cf. B1), for Latgalian (Latvia) and Asturian (Spain/Portugal) this does not

---

33 Cf. footnote 23.

34 Not listed are: Nynorsk, Catalan and Galician, which probably means that the UNESCO Atlas does not consider these to be in any danger. Low Saxon and Low German are treated as one – vulnerable – transfrontier language, which the UNESCO Atlas calls 'Low Saxon'. Its total number of speakers is estimated in the UNESCO Atlas at 4.8 Mio speakers (3 Mio in Germany, 1.8 Mio in the Netherlands). Cf. also footnote 27.

35 They are: North Frisian, Sater Frisian, Skolt Sami, Kildin Sami, Inari Sami and Jèrriais. For Lower Sorbian a status of 'severely endangered' rather than 'definitely endangered' would seem more suitable; the UNESCO Atlas mentions 'Sorbian', by which both Upper and Lower Sorbian are indicated. However, Upper Sorbian is by far the more vital of the two languages (cf. the article on Sorbian, in: Janich/Greule (2002:290) Lower Sorbian has a number of speakers that is comparable to North Frisian, a language considered to be severely endangered in the UNESCO Atlas.

36 Quite surprisingly, Latgalian is considered merely vulnerable in the UNESCO Atlas, thus placing it at a level comparable with relatively vital languages like West Frisian, Welsh and Basque. Considering the number of speakers of Latgalian as well as the (less than optimal) state of governmental recognition according to the Mercator regional dossier on Latgalian (at [http://www.mercator-research.eu/research-projects/regional-dossiers/090603.regional\\_dossier\\_latgalian\\_in\\_latvia.pdf](http://www.mercator-research.eu/research-projects/regional-dossiers/090603.regional_dossier_latgalian_in_latvia.pdf), esp. p. 7-13), a status of 'definitely endangered' would seem more realistic.

include support for lexicographical work. Since both of these are not protected by the Charter (cf. Table 1), this means that all languages in the Survey that are under the Charter's protection receive governmental support for lexicographical work.

**C2.** All languages in the Survey use an electronic corpus of primary sources. The only exception is Latgalian.<sup>37</sup>

**C3.** Languages in the Survey that use dictionary writing software (12 in total) do not use a card index system (anymore), with the exception of West Frisian, Low Saxon and Nynorsk.

**C4.** There are 10 languages in the Survey that do not use dictionary writing software. With the sole exception of Asturian, they all stem from the UK, from Germany or from the Baltic states. *Mutatis mutandis* all languages from Spain, the Scandinavian region (including Kildin Sami in Russia) and the Netherlands do use dictionary writing software.

*D. Lexicographical: Output*<sup>38</sup>

**D1.** All languages in the Survey have bilingual dictionaries and bilingual wordlists proper. Also, all languages have bilingual dictionaries of which the source language is the lesser used or non-state language.

**D2.** None of the languages in the Survey with less than 300,000 speakers (12 in total) have monolingual dictionaries, with the sole exception of Scottish Gaelic, which with 60,000 speakers does have a monolingual dictionary. Conversely, and therefore again with the exception of Scottish Gaelic, all languages that do have monolingual dictionaries are languages with 300,000+ speakers.

---

<sup>37</sup> Actually, current projects in Latgalian lexicography do not make use of a card index system either. For Latgalian, two ongoing dictionary projects were reported: one in which former dictionaries are digitalized and converted into a database format (cf. [http://www.lu.lv/filol/latgalistica/index\\_en.htm](http://www.lu.lv/filol/latgalistica/index_en.htm)), and another, private project in which Latgalian texts and audio fragments available to the editors are manually analyzed and processed (<http://www.vuordineica.lv/>). The second project was pointed out to me by Aleksey Andronov of St Petersburg State University, St Petersburg, Russia, who is personally involved in the first project only.

<sup>38</sup> Naturally funds and, as a result, manpower, constitute (the most) important factors in matters of lexicographical output. Although in the questionnaire there were no questions on these factors, one should bear in mind that differences in lexicographical output are to a large extent influenced by them.

**D3.** All languages in the Survey that have monolingual dictionaries also have bilingual dictionaries in which the source language is the dominant or official language.

**D4.** None of the languages that merely have the type of bilingual dictionary in which the source language is the lesser used or non-state language, have monolingual dictionaries.

**D5.** Languages for which no meta-lexicographical literature is reported in the Survey, include all languages from the UK (Welsh, Scottish Gaelic and Jèrriais), both languages from the Baltics (Võro and Latgalian) and all four Sami languages (Kildin, Inari, Skolt and North Sami). The two remaining languages for which no meta-lexicographical literature is reported, are Sater Frisian and Asturian.

**D6.** All languages that have dictionaries for pre-school as well as for elementary and secondary school are protected by the Charter.

Some of the results of the Survey have been linked together in order to obtain a better view on overall trends. The following section, E, contains two tables, in which several content-related questions have been brought together.

#### *E – I: Level of Diversity of Lexicographic Output*

Naturally an overview of lexicographical output in terms of the number of produced lexicographical products would render a useful impression of the state of the art of the lexicographical situation of each language. However, the questionnaires simply did not offer sufficient data for such an overview: too many informants were not able to provide the extensive bibliographical information that is necessary to produce such an overview.

A nice alternative in order to obtain an impression of the state of the art of the lexicography in lesser used language is to determine the *diversity* of lexicographical output. In order to provide such an overview, I linked together questions that are concerned with the following matters:

- monolingual dictionary (cf. Appendix, question 3.1)
- online monolingual dictionary (3.2)
- online bilingual dictionary (3.6)
- bilingual dictionary in which source language = dominant language<sup>39</sup> (3.9)

---

39 Bilingual dictionaries in which source language = lesser used language are extant for all languages in the Survey (cf. D1), thus taking these into consideration is pointless.

- monolingual wordlist (3.11)
- online monolingual wordlist (3.12)
- meta-lexicographical literature (3.31)
- pre-school/elementary school/secondary school dictionaries (0, 1, 2 or 3 points) (3.21-3.23)

Meeting each of the first 7 criteria renders 1 point, meeting the last criterion renders maximally 3 points. As a result, the maximum total of points is 10, the minimum is 0. The results are listed in Table 3.<sup>40</sup>

<b>Points</b>	<b>Language</b>	<b>No. of speakers</b>
10	<i>Basque</i>	1 Mio
10	<i>Galician</i>	2 Mio
9	<i>Catalan</i>	7-9 Mio
8	<i>West Frisian</i>	450,000
7	<i>Asturian</i>	350,000-500,000
7	<i>Nynorsk</i>	500,000
7	<i>Welsh</i>	600,000
6	<i>Lower Sorbian</i>	7,000
6	<i>Romansh</i>	60,000
6	<i>Scottish Gaelic</i>	60,000
4	<i>Friulian</i>	600,000
4	<i>Latgalian</i>	150,000-200,000
4	<i>North Frisian</i>	8,000
3	<i>Jèrriais</i>	2,600
2	<i>Low German</i>	2.6 Mio
2	<i>Low Saxon</i>	2.15 Mio
2	<i>Skolt Sami</i>	300
2	<i>Võro</i>	50,000-70,000
1	<i>Inari Sami</i>	300
1	<i>North Sami</i>	30,000
1	<i>Sater Frisian</i>	2,000
0	<i>Kildin Sami</i>	300-700

*Table 3: level of diversity of lexicographical output  
(in italics: protected by the Charter)*

<sup>40</sup> NB: a relatively low level of diversity in lexicographical output does by no means imply a low level of lexicographical output *per se*. For example, dozens of Low German, Low Saxon or North Frisian dictionaries have been compiled in the course of time, yet all of them bilingual, mostly published on paper, and merely with the lesser used language as the source language, all of which makes for a relatively low score on lexicographical diversity for these languages.

Considering a level of 6 points or higher as indicative of a (relatively) high level of diversity of lexicographical output, and 5 points or less of a (relatively) low level, leads to the following conclusions:

**E1.** Languages that are under protection of the Charter all show a (relatively) high level of diversity of lexicographical output. Asturian, again, is the exception, scoring relatively well at lexicographical diversity (7 points), although it is not under the Charter's protection.

**E2.** Lower Sorbian is the only language with less than 50,000 speakers which scores (relatively) high on lexicographical diversity.

**E3.** Lower Sorbian, with 6 points, is also the only language from Germany that scores (relatively) high on lexicographical diversity.

**E4.** None of the languages that show (relatively) low lexicographical diversity have a monolingual dictionary or wordlist, with the exception of Latgalian, which does have a monolingual wordlist.

#### *E. II: Level of Use of Modern Technology in Lexicographical Practice*

In order to give an impression of the use of modern technology in the lexicographic practice of the languages in the Survey, I devised a system comparable to the one used for Table 4. Here, the presence/absence of the following criteria was taken into account:

- online monolingual dictionary (3.2)
- online bilingual dictionary (3.6)
- online monolingual wordlist (3.12)
- online bilingual wordlist (3.16)
- use of dictionary writing software (3.33)
- use of an electronic corpus (3.32)

The criteria can be divided into two subcategories: publishing (the first 4) and production (the latter 2). Again, the presence of each criterion merits a single point. As a result, the maximum total of points is 6 and the minimum is 0. The results are in Table 4, which also shows the subdivision of the criteria.

If a score of 4 to 6 points indicates a (relatively) high level of use of technology in lexicographic practice, and 3 or less points is indicative of a (relatively) low level, Table 4 leads to the following conclusions:

<b>Language</b>	<b>Publishing</b>	<b>Production</b>	<b>Total</b>
<i>Galician</i>	4	2	6
<i>Basque</i>	3	2	5
<i>Catalan</i>	3	2	5
Friulian	2	2	4
<i>Welsh</i>	3	1	4
Asturian	3	1	4
<i>Low Saxon</i>	2	2	4
<i>Inari Sami</i>	1	2	3
<i>North Sami</i>	1	2	3
<i>Romansh</i>	1	2	3
Skolt Sami	1	2	3
Võro	2	1	3
Nynorsk	1	2	3
<i>Sater Frisian</i>	2	1	3
<i>West Frisian</i>	1	2	3
Kildin Sami	0	2	2
Jèrriais	1	1	2
<i>Lower Sorbian</i>	1	1	2
<i>Low German</i>	1	1	2
<i>Scottish Gaelic</i>	1	1	2
Latgalian	1	0	1
<i>North Frisian</i>	0	1	1

*Table 4: level of use of modern technology in lexicographical practice  
(in italics: protected by the Charter)*

**E5.** All Iberian languages show a (relatively) high level of use of technology in lexicographical practice, while all languages in Germany show a (relatively) low level.

**E6.** A total of 13 languages in the Survey score either 1 (11 languages) or 0 (North Frisian and Kildin Sami) out of a possible 4 points on online publishing. Galician is the only language with the maximum score of 4. The remaining three Iberian languages score 3 out of 4 on online publishing.

**E7.** None of the languages in Germany, in the UK or in the Baltic states use dictionary writing software. Except for Asturian, all other languages do make use of dictionary writing software

### 3 Implicational statements

At this point, I would like to comment briefly on some of the results that have been listed above, by discussing what they might imply. The results to be commented on reflect matters that I found especially striking or for which I assume that they might be of special interest to lexicographers of lesser used or non-state languages. Anyone else might have picked other topics to comment on – and everyone is encouraged to do so. The data of the questionnaires will hopefully be made available via the Euralex website, in order to put the data to the widest use possible.

First of all, it was interesting – although rather unfortunate – to see that, with the exception of Sweden Finnish, there was no response whatsoever from languages that are a dominant state language elsewhere (cf. *Response*). I suppose this will have something to do with the fact that such languages may rely on the lexicographical infrastructure of the language's 'homeland'. The number of lexicographical products especially designed for the unique situation of such languages is often quite small or even absent. There will, however, be linguistic peculiarities in such languages that should be lexicographically documented, and information on this would have enabled an interesting comparison with non-state languages. Nevertheless, participating in this Survey by filling out a rather extensive questionnaire may not have seemed relevant or worthwhile for languages that are a dominant state language elsewhere.

Another matter I should like to draw attention to is the level of use of modern technology in lexicographical practice. It seems to me that modern-day technology provides extremely helpful tools for compiling and publishing dictionaries. If nothing else, publishing costs can be cut down to the bare minimum in the case of online dictionaries and wordlists. Lexicographers of lesser used languages frequently have to make do with a rather humble amount of money – and this especially holds true for very small language communities. Online publishing therefore seems like a welcome addition to traditional, paper publishing. It is quite surprising, then, to see that among the top 10 of languages in the Survey with respect to online lexicographical publishing, not one of the 8 languages with less than 10,000 speakers can be found.

Online publishing may occur in many ways.<sup>41</sup> Uploading a PDF-file

---

41 At this point, I should like to draw attention to the Online Bibliography of Electronic Lexicography (OBELEX), set up by the Institut für Deutsche Sprache ([http://hypermedia.ids-mannheim.de/pls/lexpublic/bib\\_en.ansicht](http://hypermedia.ids-mannheim.de/pls/lexpublic/bib_en.ansicht)).

of a short wordlist is one of these, which primarily has the benefits of wide accessibility and low publishing costs over its printed equivalent. The user of such an online wordlist will, however, hardly benefit – in fact, for reasons I will not elaborate on at this time, he might even prefer the printed version. A completely different way of online lexicography, however, is compiling a scholarly dictionary online from scratch and publishing it with a keen search engine on a website. In this case the user is offered many advantages (although, admittedly, again also some disadvantages) to printed dictionaries, including easy access to the corpus on which the dictionary bases. Factors like magnitude of the language community and governmental recognition will be of influence on what medium a lexicographer chooses, since such factors for a considerable part determine the quintessential factor for any lexicographical endeavor: funds. Smaller languages that have not been officially recognized may only have modest funds at their disposal and therefore, quite understandably, be forced to use the less innovative method. Nevertheless, one must also acknowledge that not every lexicographer or lexicographical institute is equally keen on innovation.

Furthermore, it turns out that the lexicographic situation of the Iberian languages in the Survey (viz. Asturian, Basque, Catalan and Galician) is, in comparison to most other languages in the Survey, very well-developed. Both in Table 3 (on lexicographical diversity) and in Table 4 (on the use of modern technology), all four Iberian languages are in the top 6. The reason for this will be the fact that after decades of oppression by the Franco regime, in 1978 the Spanish government drew up a new constitution in which autonomous communities were created. These communities had the right to make a lesser used language co-official along with Spanish (or ‘Castilian’, as it is often called). Several language movements in Spain seized this opportunity and, as a result, the lexicography of lesser used languages benefited from these developments in the post-dictatorship period.

In contrast to the Iberian languages, Table 4 (on the use of modern technology) shows relatively low scores for all lesser used languages in Germany. This is surprising to me since, for one, Germany has a venerable lexicographic tradition, which is also expressed by a vast number of excellent dictionaries that have been compiled for the languages in Germany that are in the Survey; and also since in my experience German lexicography (that is: lexicography of modern High German and historical

lexicography in Germany) is actually quite keen on innovation.<sup>42</sup> Therefore, the relatively low scores on use of modern technology can hardly be a matter of lack of interest in innovation in the German lexicographical field. Could perhaps the fact that, except for North Frisian, the lesser used languages in Germany that are in the Survey do not receive any subsidies for setting up and maintaining a lexicographical infrastructure account for something? At any rate, one must conclude that, notwithstanding the massive lexicographical output of the lesser used languages in Germany and the high average quality of this output, the lexicographical practice of these languages is not very innovative.

I was also struck by the (relatively) low level of diversity of lexicographical output of languages with a strong lexicographical tradition such as Low Saxon and, again, North Frisian or Low German (cf. Table 3). Both Low Saxon and Low German are quite large language communities, both with over 2 Mio speakers. In the case of many smaller language communities, a lack of lexicographical diversity may, quite prosaically, be due to a low lexicographical output, which in turn is due to lack of lexicographers, which is probably ultimately caused by lack of funds. In the case of the three languages mentioned, however, the lexicographical output in itself is quite large. What may have caused the relatively low level of lexicographical diversity in this considerable output?

I suppose an important underlying reason for this may be that these languages have strong dialectal differentiation. As a result, they are at risk of having standardization issues: what dialect should become the (basis of the) language's written standard? Unless there is a dialect among them which is quite clearly the most frequently used one, or which undeniably has the highest social status, each dialect may be expected to desire becoming the (basis of the) standard. The lack of a standard makes producing a monolingual dictionary a complicated, delicate matter, as a comparable question arises: which dialect should be both the source language and the metalanguage of the dictionary?

In the case of bilingual dictionaries in which the dominant language is the source language, two strategies are possible when a commonly accepted standard for a lesser used language with strong dialectal differentiation is lacking: either the target language comprises of several or even all (main) dialects, showing some of or the complete range of dialectal variation for each entry; or the target language is a selected dialect of the language.

---

42 Cf. Popkema (2010).

In the first event, the listing of some or even all dialectal variants would involve huge amounts of time and (therefore) money, practical problems which often cannot be overcome. In the second event, again, there's the issue of what dialect should be the target language.

This problem is the more relevant since it influences overall text production in the lesser used language. In the case of lesser used languages, bilingual dictionaries in which the dominant language is the target language are often used by the language community as production dictionaries: as a rule, speakers of a lesser used language are insufficiently capable of writing it, yet do have sufficient writing skills for the dominant language. As a result, the strategy they use for writing a text in the lesser used language – their mother tongue – is looking up the word they want to write in a bilingual dictionary of which the source language is the dominant language, and then see how it is written in the lesser used language – a spell checker, if you like. Therefore, in the case of lesser used languages a bilingual dictionary in which the dominant language is the source language and the lesser used language is the target language is of vital importance for language production. Quite interestingly, in the case of speakers of a lesser used languages the strategy for writing their mother tongue is exactly the same strategy they would embrace when writing in a foreign language. At any rate, in this way the lack of a standard blocks the compiling of bilingual dictionaries in which the dominant language as the target language, which in turn blocks overall text production in the lesser used language. Thus a further diminishing of the language's social status is set into motion, placing the lesser used language in a vicious circle or sustaining the downward spiral of language decline.

Naturally, the lack of a standard in the case of languages with many sub-dialects does not mean no dictionaries are made. Quite the opposite: the existence of the sub-dialects merits the production of separate bilingual dictionaries of which each (main) dialect is the source language. This may lead to a considerable number of such bilingual dictionaries, yet not to much lexicographical diversity. And this is exactly what we see in the case of North Frisian, Low Saxon and Low German: all of them have several (main) dialects, but not a unified written standard, and all (main) dialects have their own bilingual dictionary, yet no monolingual ones or bilingual ones in which the dominant language is the source language and the lesser used language is the target language.

Conversely, there are several examples of languages that likewise have numerous sub-dialects, yet do not show the same lack of lexicographical

diversity and are at the very top of the list in Table 3: Basque, Catalan and Galician. These are languages that have strived for a written standard that supersedes the language's sub-dialects. The scope of this article does not allow me to elaborate on why or how these languages were successful in establishing a written standard while others were not – suffice to say that because of various efforts by the language movements, a written standard was established, which opened the door for a wide range of monolingual and bilingual dictionaries and wordlists. These in turn make text production in the lesser used language easier, which may further enhance the language's emancipation – a virtuous circle as opposed to the vicious circle of some of the languages without a widely accepted written standard.

In short, a (relatively) low level of diversity of lexicographical output may be caused by strong dialectal differentiation that, in turn, may obstruct the acceptance of a widely accepted written standard, which is an important factor in language emancipation.<sup>43</sup>

#### 4 Final remarks

'One of the most important issues facing humankind today is the rate at which our languages are dying. On present trends, the next century will see more than half of the world's 6800 languages become extinct, and most of these will disappear without being adequately recorded. An important first step in slowing down or reversing this process is to document the language in the form of a dictionary.'<sup>44</sup> In the light of such considerations, the lexicographer's contribution is of vital importance to documenting and preserving global cultural heritage. Therefore, it is my sincere hope that the results of the Survey – not just the results presented in this paper, but all information the combined questionnaires offer – will be of help and support to lexicographers across Europe and around the globe. Many problems lexicographers run into are of a more or less universal nature in the lexicography of lesser used languages and stories of success or failure may provide useful guidelines for lexicographical

---

43 Cf. Haugen (1966:931): "These categories [i.e. standardization and utilization in writing, ATP] suggest the path that 'underdeveloped' languages must take to become adequate instruments for a modern nation.'

44 Sarah Ogilvie (University of Cambridge) in her introductory remarks to the *Endangered Languages and Dictionaries Survey* that was recently set up (at <http://www.lucy-cav.cam.ac.uk/pages/the-college/people/sarah-ogilvie/elad1.php>). Let me take this opportunity to encourage all readers to take note of this project, and all lexicographers of lesser used languages to fill out this survey, which really is for their own benefit.

planning. Yet I also hope that the Survey may offer some form of mental support for lexicographers of severely endangered languages who are personally moved by the language's decline. For them, their work is often as tragic as it is important.

Naturally, there is a lot more to the lexicography of lesser used languages than the Survey or this paper can treat. For example, aspects of language movement and ideology are undeniably of great influence on a language's social status, on the funds the language has at its disposal and, as a result, on its lexicographical situation. Studying the interaction of language movement and lexicography would no doubt render interesting results. Furthermore, the historical background of the lexicographical tradition and infrastructure of the languages deserves proper attention. And of course, comparing the results of the Survey with the lexicography of non-European languages would be extremely interesting.

These are all matters that deserve proper attention, for which the limited space for this paper is not sufficient. I encourage every scholar that is interested in the subject matter of the Survey to visit the Euralex website, study the results of the questionnaires and use them as a basis for their own research. Hopefully, on the 15th Euralex Congress the Survey will be the subject of a paper once more. I'd like to end this Survey by thanking all lexicographers that took the time to fill out the extensive questionnaire. It is because of their efforts and input that this Survey came to be, and its results are meant to support them in their important work.

## > **References**

- Åkermark, Sia Spiliopoulou, e.a. (eds.). (2006). *International Obligations and National Debates: Minorities Around the Baltic Sea*. Mariehamn.
- Bremmer Rolf H. Jr. (2006). 'Frisian Lexicography.' In *ELL* (web-based).
- De Vries, John de (2006). 'Language Surveys.' In *ELL* (web-based).
- Brown, Keith (ed.). *Encyclopedia of Language and Linguistics*, 2nd Edition. Oxford (2006)  
(Also at: <http://www.sciencedirect.com/science/referenceworks/0780080448541>)
- Hanks, Patrick (2006). 'Lexicography: Overview.' In *ELL* (web-based).
- Hartmann, Reinhard K.K.; Gregory James (1998). *Dictionary of Lexicography*. London (1998).
- Haugen, Einar (1966). 'Dialect, Language, Nation.' In *The American Anthropologist* 68/4 (1966). 922-935.
- Hausmann, Franz. J. e.a. (eds.). (1989-1991). *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/Encyclopédie internationale de lexicographie* (3 Vols.). Berlin/New York (1989-1991) (*Handbücher zur Sprach- und Kommunikationswissenschaft* 5/1-3).

- Hawke, Andrew (2006). 'Welsh Lexicography.' In *ELL* (web-based).
- Hinderling, Robert; Ludwig M.Eichinger (1996). *Handbuch der mitteleuropäischen Sprachminderheiten*. Tübingen.
- Holtus, Günter; Michael Metzeltin; Christian Schmitt (eds.). (1998-2005). *Lexikon der romanischen Linguistik* (8 Vols.). Tübingen.
- Inoue, Miyak (2006). 'Standardization.' In *ELL* (web-based).
- Janich, Nina; Albrecht Greule (2002). (eds.). *Sprachkulturen in Europa. Ein internationales Handbuch*. Tübingen.
- Kulbrandstad, Lars A.; Olav Veka (2006). 'Scandinavian Lexicography.' In *ELL* (web-based).
- LLR. See Holtus, e.a.
- Niebaum, Hermann (2008a). 'Het Oostnederlandse taallandschap tot het begin van de 19e eeuw.' In Jurjen van der Kooi, e.a. (eds.), *Handboek der Nedersaksische Taal- en Letterkunde*. Assen (2008), 52-64.
- Niebaum, Hermann (2008b). 'Het Nederduits.' In Jurjen van der Kooi, e.a. (eds.), *Handboek der Nedersaksische Taal- en Letterkunde*. Assen (2008). 430-447.
- Popkema, Anne T. (2010). 'Eine Perspektive der altfriesischen Lexikographie? Zu einem Online-Belegwörterbuch des Altfrisischen.' In Piter Boersma e.a. (eds.), *Philologia Frisica anno 2008*. Leeuwarden (expected in 2010).
- Saurí Colomer, Roser (2006). 'Spain: Lexicography in Iberian Languages.' In *ELL* (web-based).
- Simpson, J. (1993). 'Making Dictionaries.' In Michael Walsh/Colin Yallop (eds.), *Language and Culture in Aboriginal Australia*. Canberra (1993). 123-144.
- Svensén, Bo (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge.

## > **Web-based (April 2010)**

ELL

Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd Edition. Oxford (2006).

<http://www.sciencedirect.com/science/referenceworks/9780080448541>

Endangered Languages and Dictionaries Survey

<http://www.lucy-cav.cam.ac.uk/pages/the-college/people/sarah-ogilvie/elad1.php>

Ethnologue Languages of the World (16th Edition)

<http://www.ethnologue.com/web.asp>

European Charter for Regional or Minority Languages

<http://conventions.coe.int/treaty/en/Treaties/Html/148.htm>

European Charter for Regional or Minority Languages: Explanatory Report

<http://conventions.coe.int/treaty/en/Reports/Html/148.htm>

Euralex

<http://www.euralex2010.eu/>

Mercator Facts and Figures

<http://www.mercator-research.eu/minority-languages/facts-figures>

Mercator Regional Dossiers

<http://www.mercator-research.eu/research-projects/regional-dossiers>)

Mercator Database of Experts

<http://www.mercator-research.eu/minority-languages/database%20of%20experts>

OBELEX

[http://hypermedia.ids-mannheim.de/pls/lexpublic/bib\\_en.ansicht](http://hypermedia.ids-mannheim.de/pls/lexpublic/bib_en.ansicht)

UNESCO Atlas of Languages in Danger

<http://www.unesco.org/culture/ich/index.php?pg=00206>

UNESCO (2003)

UNESCO, 'Language Vitality and Endangerment'. UNESCO Ad Hoc Expert Group on Endangered Languages. Document submitted to the *International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages* Paris, 10–12 March 2003

<http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf>

> **Appendix**

**Questionnaire concerning lexicography of European lesser used languages**

*Introduction*

The Fryske Akademy in Leeuwarden, The Netherlands does research on Frisian and Friesland. Over the years the Fryske Akademy has published a wide range of Frisian dictionaries. The Mercator Research Centre at the Fryske Akademy studies the position of lesser-used languages in education, gives information on national and regional education systems and provides the latest statistics regarding lesser-used languages in education in the European Union.

The European Association for Lexicography (Euralex) holds biennial congresses, attended by several hundred people, where refereed papers are presented on a large variety of topics relevant to its members' interests. The Fryske Akademy will host the next congress, which will be held in Leeuwarden/Ljouwert, the Netherlands, from 6 – 10 July 2010. See the congress web site for more information. One of the special features of the 2010 conference is its focus on the lexicography of lesser used non-state languages. In preparation of the conference we would like to learn more about the state of the art of the lexicography of the individual lesser used languages in Europe and also about their social and linguistic situation. To that end we have compiled the following survey that we hope you will complete. The results of the survey will be presented at the Euralex conference in 2010.

**1 Contact data of the informant**

- 1.1 last name:
- 1.2 first name:
- 1.3 address:
- 1.4 city:
- 1.5 country:
- 1.6 e-mail:
- 1.7 affiliation:
- 1.8 mailing-list:
- 1.9 website:

**2 Social and linguistic situation of the lesser used language**

- 2.1 what language are you/is your institute dealing with?  
local name:  
English name:

- 2.2 in what region(s) and country/countries is the language spoken?  
 .....
- 2.3 how many speakers does the language have?  
 .....
- 2.4 to which larger national language is the language linguistically related?  
 .....
- 2.5 has the language been recognized by the national government?  
 no  yes  
 – if yes, in what way?  
 .....
- 2.6 does the language receive support to survive?  
 no  
 yes, from  
 .....
- 2.7 does this include support for lexicographical work?  no  yes
- 2.8 do the national authorities consider lexicography as a means for language maintenance?  no  yes
- 2.9 do the local authorities consider lexicography as a means for language maintenance?  no  yes
- 2.10 does the language have an organic place in  
 – education  no  yes  
 – the media  no  yes  
 – other, namely  
 .....
- 2.11 does the language have an official spelling?  no  yes
- 2.12 if yes, who determines the spelling rules?  
 .....
- 2.13 are there grammars of the language?  no  yes
- 2.14 if yes, do the dictionaries (if any) follow the available grammars?  
 no  yes
- 2.15 do you consider the lexical distance to the official language  
 – considerable?  no  yes  
 – marginal?  no  yes
- 2.16 do you consider the syntactic distance to the official language  
 – considerable?  no  yes  
 – marginal?  no  yes
- 2.17 is there a tendency among educated speakers of the language to use words and/or constructions that most clearly show the differences with the dominant language (distancing)?  no  yes

### 3 Lexicographic situation of the language

- 3.1 are there monolingual dictionaries of the language?  no  yes

- if yes, please give the title(s) and the year(s) of publication  
 .....
- 3.2 are they on-line?  no  yes  
 – if yes, please give the URL's:  
 .....
- 3.3 are they on paper?  no  yes  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....
- 3.4 are they on CD-ROM?  no  yes  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....
- 3.5 are there bilingual dictionaries of the language?  no  yes  
 – if yes, please give the title(s) and the year(s) of publication:  
 .....
- 3.6 are they on-line?  no  yes  
 – if yes, please give the URL's:  
 .....
- 3.7 are they on paper?  no  yes  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....
- 3.8 are they on CD-ROM?  no  yes  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....
- 3.9 is the source language of the dictionaries  
 – the lesser used language?  no  yes  
 – the standard/dominant language?  no  yes
- 3.10 is the target/explanatory language a widely used language like  
 English in order to reach an international audience of linguists  
 and/or lexicographers?  no  yes
- 3.11 are there monolingual wordlists of the language?  no  yes  
 – if yes, please give the title(s) and the year(s) of publication:  
 .....
- 3.12 are they on-line wordlists?  no  yes  
 – if yes, please give the URL's:  
 .....
- 3.13 are they paper wordlists?  no  yes  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....

- 3.14 are they on CD-ROM?  no  yes  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....
- 3.15 are there bilingual wordlists of the language?  no  yes  
 – if yes, please give the title(s) and the year(s) of publication:  
 .....
- 3.16 are they on-line wordlists?  no  yes  
 – if yes, please give the URL's:  
 .....
- 3.17 are they paper wordlists?  no  yes  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....
- 3.18 are they on CD-ROM?  no  yes  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....
- 3.19 is the source language of the wordlists  
 – the lesser used language?  no  yes  
 – the standard/dominant language?  no  yes
- 3.20 what is the goal/background of the dictionaries/wordlists?  
 mainly scientific?  
 mainly practical?  
 mainly educational?  
 mainly touristic  
 other? Namely:  
 .....
- 3.21 are there pre-school picture dictionaries?  no  yes  
 – if yes, please give the title(s) and the year(s) of publication:  
 .....
- 3.22 are there dictionaries for elementary schools?  no  yes  
 – if yes, please give the title(s) and the year(s) of publication:  
 .....  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....
- 3.23 are there dictionaries for secondary schools?  no  yes  
 – if yes, please give the title(s) and the year(s) of publication:  
 .....  
 – if yes, how many copies were  
 – printed? .....  
 – sold? .....

- 3.24 are the dictionaries/wordlists a result of  
 – a private initiative?  no  yes  
 – an institutional initiative?  no  yes
- 3.25 are the dictionaries published  
 – privately?  no  yes  
 – by professional publishers?  no  yes  
 – on-line?  no  yes  
 – if yes, please give the URL's:  
 .....
- 3.26 are there any subsidies to facilitate the publication of dictionaries?  
 no  yes
- 3.27 are there any subsidies to help setting up and maintaining a  
 lexicographical infrastructure?  no  yes
- 3.28 have there been studies concerning the dictionary's target group  
 and its needs?  no  yes
- 3.29 do you experience difficulties distributing lexicographical  
 products?  no  yes  
 – if yes, please specify:  
 .....
- 3.30 is the lexicography of the language embedded in other linguistic  
 research?  no  yes
- 3.31 are there (theoretical) publications on the lexicographical practice?  
 no  yes
- 3.32 do you work with  
 – a card-index corpus?  no  yes  
 – an electronic corpus?  no  yes
- 3.33 do you work with a dictionary compilation program?  
 no  
 yes, a custom made program  
 yes, a commercial program, namely:  
 .....
- 3.34 do you consult fellow lexicographers of lesser used languages?  
 no  yes
- 3.35 would you consider it useful to have an on-line forum for  
 lexicographers of lesser used languages?  no  yes