The image of the monolingual dictionary across Europe: Results of the European survey of dictionary use and culture

Abstract

The article presents the results of a survey on dictionary use in Europe, focusing on general monolingual dictionaries. The survey is the broadest survey of dictionary use to date, covering close to 10,000 dictionary users (and non-users) in nearly thirty countries. Our survey covers varied user groups, going beyond the students and translators who have tended to dominate such studies thus far. The survey was delivered via an online survey platform, in language versions specific to each target country. It was completed by 9,562 respondents, over 300 respondents per country on average. The survey consisted of the general section, which was translated and presented to all participants, as well as country-specific sections for a subset of 11 countries, which were drafted by collaborators at the national level. The present report covers the general section.

1 Introduction

Research into dictionary use has become increasingly important in recent years. In contrast to 15 years ago, new findings in this area are presented every year, e.g. at every Euralex or eLex conference. These studies range from questionnaire or log file studies to smaller-scale studies focussing on eye tracking, usability, or other aspects of dictionary use measurable in a lab. For an overview of different studies, see Atkins (1998); Welker (2010); Lew (2011); Töpel (2014); Lew (2015a); Müller-Spitzer et al. (2018). Influential individual contributions include: Béjoint (1989); Benbow et al. (1990); Atkins and Varantola (1997); Nesi (2000); Tono (2000 ; 2001); Lew (2002); Jopling (2003); Boonmoh and Nesi (2008); Dziemianko (2011); Frankenberg-Garcia (2011); Nesi and Tan (2011); Dziemianko (2012); studies in Müller-Spitzer (2014); Müller-Spitzer et al. (2015); Dziemianko (2016). Although the field has made impressive progress, studies have tended to focus on students and language professionals, rather than on the more general dictionary-using public (cf. Lew 2015b: 234). This is largely due to the fact that students form, as it were, a captive audience for academic research, and, for large-scale surveys, distributional channels such as mailing lists are used that are dominated by language professionals and students. This was, for example, the case in the large-scale questionnaire studies reported in Müller-Spitzer (2014). Moreover, monolingual dictionaries have rarely been in the spotlight of interest (Klosa et al. 2014 is one exception). In an effort to redress this imbalance, the study reported here focused on monolingual dictionaries and tried to reach a broad sample by carefully disseminating a survey via multiple channels, and in multiple language versions (see Section 3.1). To help with the translation of the questionnaire and its dissemination, the core group approached local researchers/lexicographers in individual countries (hereafter: partners), identified through the European Network of e-Lexicography (ENeL; EU COST Action IS1305),[1] or using existing contacts of the members of the co-ordination group. Fifty-eight researchers from 29 different countries became involved;[2] however, in the results presented in this paper, three countries have been excluded, as they did not meet the minimum threshold of 100 completed questionnaires (see Section 4.1).

The remainder of the paper is structured as follows: in the next section, the situation of monolingual dictionaries in Europe is described (Section 2). In Section 3, we introduce the survey, setting out its general principles and aims, as well as the implementation and the process of translation. Section 4 presents the results, followed by discussion in Section 5. The paper ends with concluding remarks.

2 Monolingual dictionaries in Europe

The European lexicographic landscape is very diverse, with different languages having different lexicographic traditions and available resources. This has been observed in ENeL meetings[3] and recorded in studies such as the META.NET white paper series (Key Results and Cross-Language Comparison 2012). In order to get a good understanding of similarities and differences in participating countries, we asked our partners to provide us with short overviews of existing monolingual dictionary situations in their respective countries, most of which are made available as Supplementary Online Material.[4] This information was also essential in ensuring the correct interpretation of the results of our survey. In the remainder of this section, we rely heavily on the details reported by our partners in participating countries.

As far as the number of monolingual dictionaries available on the market is concerned, most languages have more than one. For some languages, there are many monolingual dictionaries on the market; this is especially true for languages with a high number of native speakers, such as English, French, Italian, Spanish and German, but also for languages such as Danish and Greek. There are, however, a few languages or countries where only one general monolingual dictionary (in different editions or with derivatives that are smaller in size) currently exists; these include Estonian, Finnish, Georgian, and Slovenian. The Basque Country, Belgium, and Ireland are special cases: in the Basque Country, bilingual dictionaries have equal status to, if not more important status than, monolingual dictionaries; in Belgium, the dictionaries that are used are published in France (for French) and in the Netherlands (for Dutch); and in Ireland, while Irish is the official language, only a minority of people speak it, so English dictionaries are mainly relevant.

For nearly all languages, with the exception of Finnish and Hebrew (in Israel), spelling or orthographic dictionaries play an important role in monolingual lexicography. In fact, in countries such as Croatia, the Czech Republic, Germany, Portugal, Slovenia, and Sweden, they are the ones considered to represent the language norm (official or otherwise). On the other hand, in countries such as Austria, Finland, Greece, Italy, Norway, Poland, the UK, Ireland, and a few others, this role is still taken by the general monolingual dictionaries.

Significant differences can also be observed in terms of formats in which monolingual dictionaries are available. In countries such as Austria, Croatia, France, Georgia, Greece, Israel, Italy, Macedonia, Portugal, Romania, and Serbia, paper dictionaries still dominate; few digital (primarily online) dictionaries are available, but their number is on the increase. However, the partners usually noted that it is the digital dictionaries that are probably more popular among the users — which is something to be examined in the survey. Interestingly, in Serbia there was no digital version of a general monolingual dictionary available at the time when the survey was conducted.

For most languages, online dictionaries are merely digitized versions of printed counterparts, a well-known fact often pointed out in the lexicographic literature (e.g. Rundell 2015: 305). In countries such as Romania and Serbia, the current focus of monolingual lexicography is more on retro-digitizing existing printed monolingual dictionaries. In contrast, in the Netherlands and Poland, born-digital dictionaries are already available, and similar initiatives, although still in early stages, can be observed in the Czech Republic, Croatia, and Slovenia.

In the majority of the countries participating in the survey, monolingual dictionaries are published solely or mainly by public institutions funded by the government. This tends to be connected with the small number of native speakers (around 10 million or fewer), and, in consequence, small buyer markets (e.g. in the Czech Republic, Estonia, Finland, Georgia, Serbia, Slovenia and Sweden); however, there are notable exceptions, such as Poland and Spain.

In Croatia and Norway, even if commercial publishers compile a monolingual dictionary, they are partly or entirely funded with public money. Commercial publishers dominate over public institutions in Greece, Germany, France, Israel, Italy, Portugal, and the UK. In Denmark, Greece, Ireland, and the Netherlands, where publicly funded dictionaries are dominant, commercial publishers still play an important role.

The type of funding also dictates the business model used for providing online dictionaries. Therefore, publicly-funded online dictionaries are usually free, whereas online dictionaries published by commercial publishers are normally available for a one-off fee or an annual subscription. Certain commercial publishers in countries such as Croatia and Italy use a model where you need to buy a print version of the dictionary in order to get access to the online version. The same model has been used in Slovenia for a publicly-funded monolingual dictionary.

Another piece of information obtained in these overviews was whether dictionaries are used in schools, in particular whether they are included in the curriculum. It turns out that in twelve out of twenty-six countries (Austria, Belgium, Denmark, Estonia, France, Greece, Norway, Portugal, Slovenia, Spain, and the United Kingdom) dictionaries and dictionary skills are specifically mentioned in the curriculum, while in the remaining eleven countries they are not. Most countries also feature school dictionaries, with Estonia, Serbia, Slovenia, and Sweden being exceptions in this regard.

Nearly all the countries have one other common denominator related to monolingual dictionaries: the field of dictionary use research is very poorly developed or non-existent. The UK, and to a lesser extent Denmark and Germany, are notable exceptions. For example, Germany now has a competence centre for research into dictionary use, based at the Institut für Deutsche Sprache (IDS) in Mannheim. Still, it is encouraging that many partners reported that research into dictionary use has been gaining strength in recent years. The survey presented in this paper will hopefully stimulate such research in participating countries.

3 The present survey

3.1 *General principles and aims*

The idea behind our survey was to attempt to include as many EU countries as possible, plus a few non-member countries with close cultural ties to Europe and with active members in the ENeL network. We aimed at a large number of participants, so well-thought-out promotion was of the essence. In order not to discourage participants from completing the survey, it was essential that the survey did not take a lot of their time. A natural consequence of this assumption was that we needed a fairly small number of carefully selected questions.

Another requirement of our survey was to try to cover participants of varied age and background, including those who chose not to use dictionaries, since we wanted to learn their views on dictionaries, what it was that stopped them from using dictionaries, and what it would take for them to start using them (again).

The most central research aim was to probe the usage patterns of, and attitudes towards, general monolingual native language dictionaries. At the same time, we kept the participants' options open as far as the medium of the dictionary was concerned, acknowledging that there might be substantial differences across the countries in this respect: some might have largely adopted the newer digital formats, while others would still embrace print (see Section 2).

In formulating the survey items, we were aiming to make it possible to compare countries, as well as measure the effects of participant age and background. With this assumption, the

items had to make sense to a broad audience across a diverse spectrum of languages and cultures. In a similar vein, we were aware of the gap in previous research relating to the fact that surveys have tended to capture participants as dictionary users, but virtually no surveys have sought responses from people who did *not* use dictionaries. We did want a representation of the latter group in our survey, challenging as this might be. A natural consequence of this assumption was that as well as not being too long, the survey could not be too long, and should not assume detailed knowledge about dictionaries on the part of the participants. The features of brevity and generality should also go a long way towards making it possible to reuse the survey in the future in other countries (also outside Europe), and possibly for diachronic comparisons.

3.2 *Structure and implementation*

The core group consisting of Iztok Kosem, Robert Lew, Carolin Müller-Spitzer, Maria Ribeiro Silveira, and Sascha Wolfer drafted 13 questions that formed the *general part* of the survey. These questions were accompanied by 11 questions eliciting personal data from the participants. Henceforth, we refer to the latter items as *meta variables*. In constructing the survey, the group consulted a number of experts, including an expert on social survey methods. It was also piloted among students at the University of Mannheim and University of Ljubljana.

The core group distributed the general part along with the meta variable items to partners in the participating countries. They were responsible for translating the questions into their local language(s). The translation process included discussions with the core group, in order to avoid losing too much in translation and to make sure the answers obtained in different countries would be comparable.

In addition, partners were given an option to contribute local questions in case they wanted to pursue a research question that was not covered by the general questions. We asked the local representatives to restrict their local parts to a maximum of five items, so as to keep the total length manageable. The local questions included topics such as types of information most often consulted in a dictionary, potential improvements to a general monolingual dictionary in a particular country, use of dialectal dictionaries and bilingual dictionaries, experience with and views on user contributions in dictionary compilation, and use of dictionaries in educational settings. The local questions were presented to participants from their respective countries only, but we asked for an English translation of the local questions to enable the core group to give feedback on the questions. In this article, we do not deal with any of the local questions. The partners in the participating countries are in charge of analysing this data and publishing the corresponding results.

For dissemination, we used a variety of channels, including mailing lists, institutional websites, social media such as Facebook and Twitter, and personal contacts. To simplify the dissemination of the survey, we created a single webpage, to which all potential participants could be directed by the researchers in the participating countries. On this page, visitors were presented with a very short welcome message and a list of countries and languages. When they selected a language from this list, they were taken to the respective language version of the questionnaire. The questionnaire itself was implemented in the commercial online survey system *QuestBack Unipark*. The different language versions were activated by means of a URL parameter encoding the language by a number. The online survey was active between the 8th of May and 9th of July 2017. A static dump of the complete general section of the Survey is supplied as **Supplementary Online Material**.

On the first page of the survey, participants were presented with a welcome message, a declaration of consent, and an e-mail address to contact Sascha Wolfer with any questions they

might have. By clicking on the *Continue* button on the welcome page, participants declared that their participation was voluntary, and that they understood that they could omit any questions that they did not wish to answer. Throughout the questionnaire, this was implemented by using 'skippable obligatory questions', whereby a confirmation dialogue would appear when an item was left blank, and participants were able to confirm that it was their intention to skip an item. In the declaration of consent, participants were assured that their responses would only be reported in aggregated form, and that only the researchers would have access to their individual responses.

The first question of the survey was the only one that concerned dictionaries in general. After the first question, a short information text was presented stating that the remainder of the survey would 'focus on general monolingual dictionaries of [language]', where the placeholder [language] was filled in with the language of the questionnaire. To address all potential types of participants, dictionary users and non-users, we characterised general monolingual dictionaries as simply as possible: 'A general monolingual dictionary of [language] describes [language] words using [language] explanations. There is no other language involved. In such a dictionary you can find most commonly used words.' After this short information, the survey continued without any intervening information screens.

At several points throughout this text and the figures presented therein, we use abbreviated versions of questions and/or response options. Full original text of the general questions, the meta-variable items, and all the response options are supplied as **Supplementary Online Material**.

4 Results

4.1 *Overview of the sample*

Altogether, responses from 9,099 participants were included in the analysis. The following meta variables were collected: year of birth, years of formal education, gender, occupational status, whether a participant was a language teacher, whether a participant had completed, or was working towards, a university degree in which language or linguistics was a major component, whether a participant's job had a strong focus on language, whether a participant had a special interest in languages and/or their native language. Due to the nature of the study, we had to rely on the self-reports of the participants.

Forty-eight participants (0.53%) did not enter their year of birth. Furthermore, to control for some of the most implausible answers, we treated all birth years prior to 1901 as NA values ('not available'). This affected 45 individuals (0.50%). All other birth years were transformed into ages. The mean age of the participants was 38.8 years (median: 37). Half of all participants were between 26 and 49 years of age (inter-quartile range). For further analyses, we calculated six percentile-based age groups (i.e. group boundaries were calculated from the data in a way that the groups were of roughly equal size): 1,753 participants (19.5%) were 24 years old or younger, 1,470 (16.3%) were between 25 and 30 years old, 1,403 (15.6%) were between 31 and 37 years old, 1,396 (15.5%) between 38 and 44, 1,500 (16.7%) between 45 and 54, and 1,484 participants (16.5%) were 55 years old or older.

We chose to ask for the years of formal education because we did not want to confuse participants with a complex question about their highest educational qualification that would make allowances for the idiosyncrasies of educational systems across Europe. Forty-one participants (0.45%) did not report the number of years of formal education they had received. We programmed in a validation rule that triggered an error message whenever the years of formal education exceeded the age of the participant. To eliminate further implausible answers,

we treated all entries above 50 years as NA values. This affected 16 individuals (0.18%). The mean number of years of education was 17.7 years (median: 18). The inter-quartile range of years of formal education was 16 to 20. We also created educational groups analogous to age groups. Due to the distribution of educational years, we had to create four groups. The group boundaries and sizes were as follows. 16 years and less: 3,333 (35.8%), 17 and 18 years: 2,359 (26.1%), 19 and 20 years: 1,681 (18.6%), 21 years and over: 1,769 (19.6%).

6,470 (69.1%) of the participants were female and 2,718 (28.9%) were male. 179 people (2.0%) did not want to give information about their gender. Presumably, this figure is comparably high because, in this case, we provided an explicit 'I don't want to answer' option (in case a participant did not identify with either of the binary gender categories: *female* or *male*). Six (0.07%) people did not provide an answer at all.

4,367 people (48.0%) were employees in the private or public sector, 1,872 (20.6%) participants were undergraduate or Master's students, and 781 (8.6%) were self-employed. 595 (6.5%) participants were Ph.D. students, 480 (5.3%) were retired. 252 (2.8%) participants were pupils in secondary education, and 202 (2.2%) were unemployed at the time of completing the survey. A further 58 (0.6%) were homemakers and 44 (0.5%) were trainees or apprentices. 363 people (4.0%) chose the option 'Other'. 85 participants (1.0%) did not supply information on their occupational status.

When we look at the status of language within our sample, we see that 2,457 participants (27.0%) were language teachers. 6,634 (72.9%) were not, and 8 (0.1%) did not provide an answer. 4,864 participants (53.5%) were studying or had studied a subject in which language or linguistics was a major component, 4,220 (46.4%) were not and had not, 15 (0.16%) did not provide an answer. The results for the questions about the status of language in the participants' jobs and whether they had a special interest in language are summarised in **Table 1**.

Table 1. Distribution of responses to questions about job focus on language and special interest in language.

Does your job have a focus on language (either native or foreign)?					
Strongly agree	Agree	Neither … nor	Disagree	Strongly disagree	Missing
4205	1947	1195	1412	306	34
46.2%	21.4%	13.1%	15.5%	3.4%	0.37%
Would you say that you have a special interest in language?					
Strongly agree	Agree	Neither … nor	Disagree	Strongly disagree	Missing
5067	2721	984	273	45	9
55.7%	29.9%	10.8%	3.0%	0.49%	0.10%

When we look at the overall distribution of these participant variables, we see that the participants distribute over a wide range of ages. Women are over-represented in our sample, but there are enough male participants to conduct contrastive analyses where this seems appropriate. When it comes to the involvement with language-related or linguistic issues, we see that the survey tended to attract people that are professionally involved or at least interested in language. For example, current or former students of linguistics or another subject in which language is a major component are the majority in our sample. The item about special interest in language is also a good indicator of this. Roughly 86% of the participants agreed (including 'strongly agreed') that they had a special interest in language. That is not at all surprising because this is a characteristic which would make people likely to take part in a study investigating dictionary use and culture. This is a type of sampling bias and needs to be borne in mind when interpreting the results. The number of participants not interested in language at

all (or not having a language-related job) is probably too small to do contrastive analyses along this dimension.

When it comes to the distribution of participants over participating countries, **Table 2** gives an overview of the number and share of the participants' home countries/areas. As can be expected in a relatively uncontrolled data collection setting, the participants distributed quite unevenly across the countries. To avoid even more extremely skewed distributions, we excluded from the final dataset any countries that did not manage to collect at least 100 participants. As a result, all participants from Hungary (47 participants), Iceland (48 participants), and Latvia (94 participants) were excluded. However, the ratio between the country with the most (Romania) and least (Sweden) participants is still 878:112 = 7.84, i.e. there are nearly eight times as many participants from Romania than there are from Sweden.[5]

Table 2. Distribution of participants over countries/areas.

Country	*n*	Share
Romania	878	9.65%
Greece	829	9.11%
Poland	649	7.13%
Slovenia	619	6.80%
Croatia	516	5.67%
Georgia	507	5.57%
Germany	479	5.26%
Estonia	467	5.13%
Norway	420	4.62%
Denmark	405	4.45%
Portugal	349	3.84%
Finland	298	3.28%
Spain	297	3.26%
Serbia	293	3.22%
Belgium	286	3.13%
Italy	285	3.13%
Austria	268	2.95%
France	238	2.62%
Macedonia	181	1.99%
UK & Ireland	169	1.80%
Czech Rep.	146	1.60%
Israel	146	1.60%
Basque Cnt.	132	1.45%
Netherlands	130	1.43%
Sweden	112	1.23%

4.2 *Using dictionaries*

First, we investigated how often, on which devices, and in which situations monolingual dictionaries were used. We started the survey with a question regarding any types of dictionaries (Question 1) and then restricted the survey to monolingual dictionaries only (Question 2). The difference was, as noted above, explained on an extra screen. This allowed us to compare the frequency of use of monolingual dictionaries versus any dictionary type. **Figure 1** gives an overview of the responses. Note that we chose to present answer options that deviate from the (maybe more traditional) options 'once per day/week/month…'. We did this because we wanted

participants to try and mentally recall their last instance of dictionary use, which should be more reliable than giving a rough estimate of regularity that is implied by the traditional form. Roughly half of the participants (51.8%) used a dictionary 'yesterday or today', but only a third (33.2%) used a *monolingual* dictionary 'yesterday or today'. The answers for the next frequency category 'within the last week' are equal (25% each). For even longer time periods implying less frequent dictionary use, monolingual dictionaries dominate over dictionaries in general. It is not surprising that monolingual dictionaries are used less frequently than dictionaries in general, because the former are a subset of the latter. 501 participants (5.5%) never used a monolingual dictionary. In contrast, only 104 participants (1.1%) never used a dictionary at all. We see that the majority of our participants (58%) used monolingual dictionaries at least on a weekly basis and 94% of all participants used a monolingual dictionary at least once. This implies they are in a position to give informed answers to the remaining survey questions, at least based on their self-report.

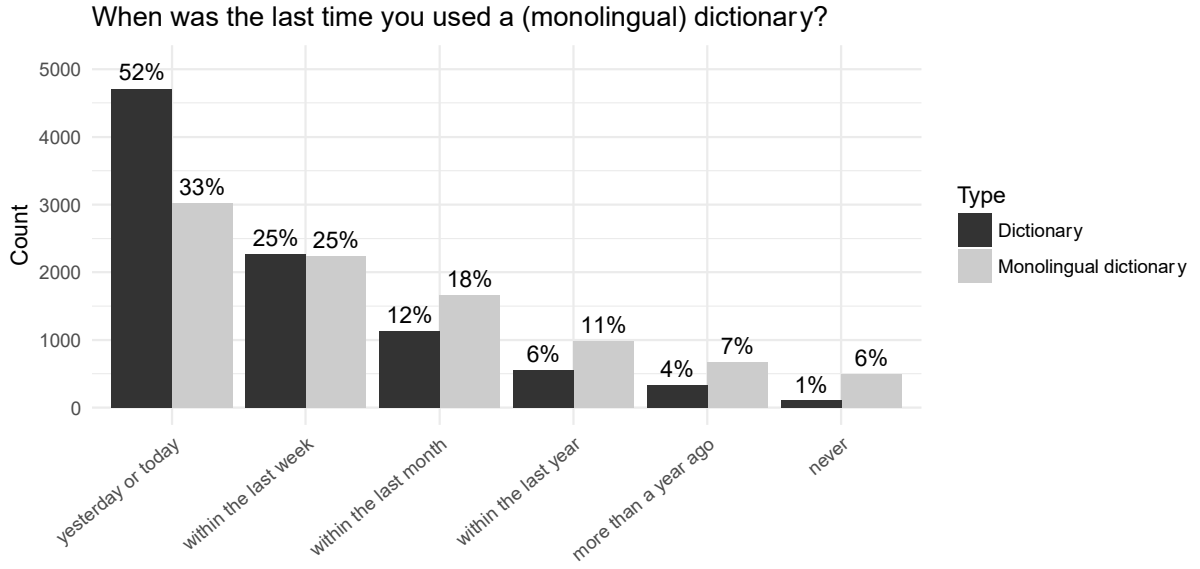

Figure 1. Responses to Question 1 (When was the last time you used a dictionary?) and Question 2 (When was the last time you used a monolingual dictionary?). Participants had to select one option for each of the two questions.

The next three questions were presented only to those participants who used a monolingual dictionary at least once (this was handled by the flow control logic of the survey software). In Question 4, participants were asked to check all applicable answers concerning the format of monolingual dictionary they *used*. In Question 5, in contrast, they had to decide which format they *preferred*. **Figure 2** shows the distribution of answers to both questions. We see that the preferences of participants lined up with their actual use. However, although 13.7% of the participants used dictionaries on tablets, only 2.4% actually preferred them over the other formats. At the other side of the spectrum, the computer (this included desktop and laptop computers) was used and preferred by most participants, nearly twice the rate of print dictionaries in terms of preference. When comparing mobile devices (smartphones and tablets), smartphones outperformed tablets considerably, both in terms of use and preference.

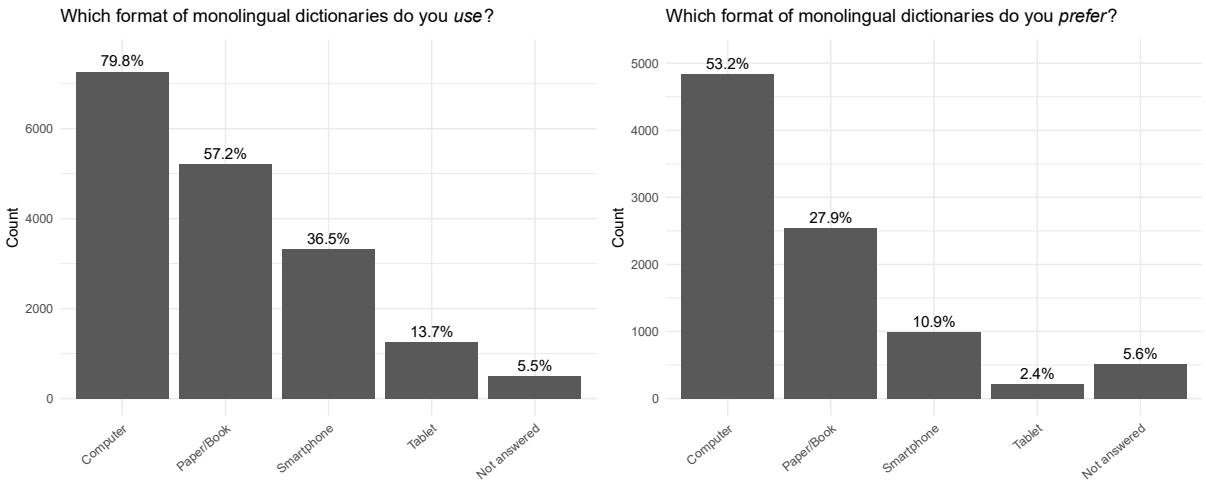

Figure 2: Distribution of responses to Question 4 (left) and 5 (right). English versions of the questions are given as titles in the images. Percentages in the left panel sum up to more than 100% because multiple responses were allowed. Participants had to select one option only for Question 5, and so here percentages sum up to 100%. 501 participants indicated that they never used a monolingual dictionary of their language. These participants did not answer questions 4 and 5. No participant skipped Question 4, 12 participants skipped Question 5.

The distribution of formats of monolingual dictionaries used depended on the age of the participants. The older the participants, the more likely they were to use books (r = 0.17), less likely to use computers (r = -0.09), more likely to use tablets (r = 0.05) and less likely to use smartphones to access monolingual dictionaries (r = -0.23; all Pearson correlation coefficients are highly significant, $p < 0.0001$). For the youngest age group (24 years and younger), print dictionaries were roughly equally likely to be used (48.8%) as smartphone-based dictionaries (47.0%), though neither nearly as likely as computer-based dictionaries (74.8%). Thus, younger participants clearly preferred the computer for accessing monolingual dictionaries. In the oldest age group, by contrast, print dictionaries were on a par (72.3%) with dictionaries on a computer (72.9%), with smartphones being used much less often (19.2%).

It is quite clear, both from the overall analysis and from the group-based analysis, that computers were the favourite and most used device on which dictionaries were accessed. We can also combine this question with another meta variable we did not introduce above: the devices our participants used on a daily basis. Participants were asked to choose one or more of the following devices: desktop computer, laptop, tablet, and smartphone. The answers to this question were not surprising: most participants claimed to use smartphones on a daily basis (82.5%), followed by laptops (72.4%), desktop computers (52.6%), and tablets (24.2%). When responses for laptops and desktop computers are conflated, then almost all participants reported using a computer of either type (94.5%). One notable observation that arises from combining these figures with those for used and preferred formats for monolingual dictionaries is that although many people were using smartphones, they were not using them that much for accessing monolingual dictionaries. Using our data, we can also analyse this directly (see **Figure 3**).

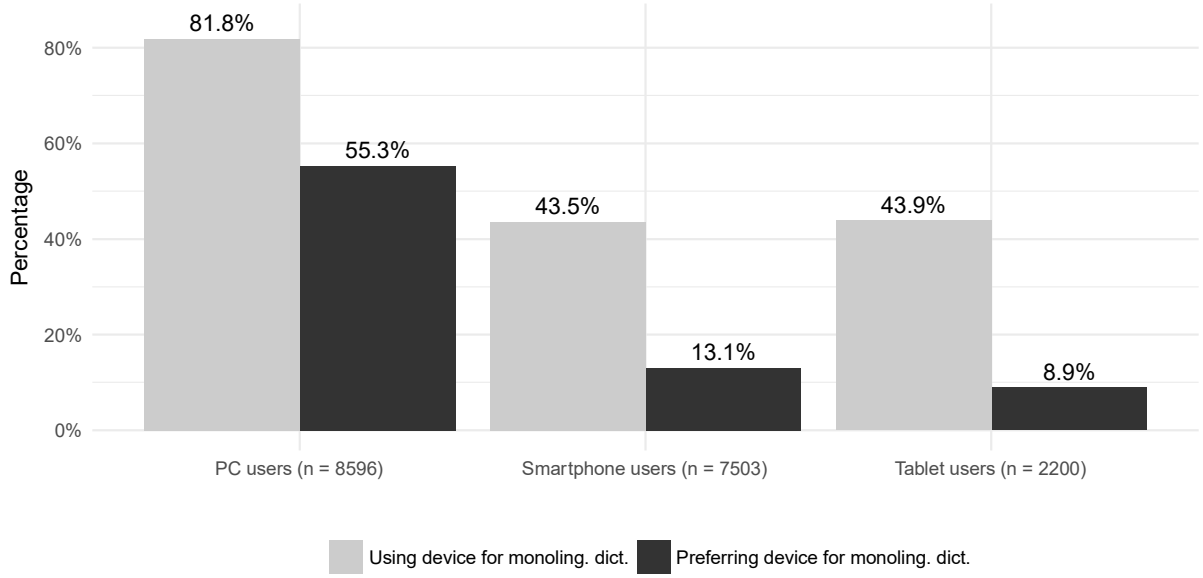

Figure 3: Three user groups (x-axis) using the respective device on a daily basis and how many of those used (light bars) and preferred (dark bars) the respective device for accessing monolingual dictionaries (y-axis). Group sizes are given in the x-axis labels. The groups are not mutually exclusive.

As many as 8,596 people in our sample used a computer on a daily basis and 7,503 people were daily smartphone users. Obviously, the two groups are not mutually exclusive. Out of all computer users, 81.8% were also using their computer to access monolingual dictionaries. The majority of computer users (55.3%) preferred to access monolingual dictionaries on a computer. For smartphone users, the picture is dramatically different: only 43.5% of them used their smartphones to access monolingual dictionaries (they are on a par with tablet users in this regard) and only as few as 13.1% chose them as their preferred device for monolingual dictionary access.

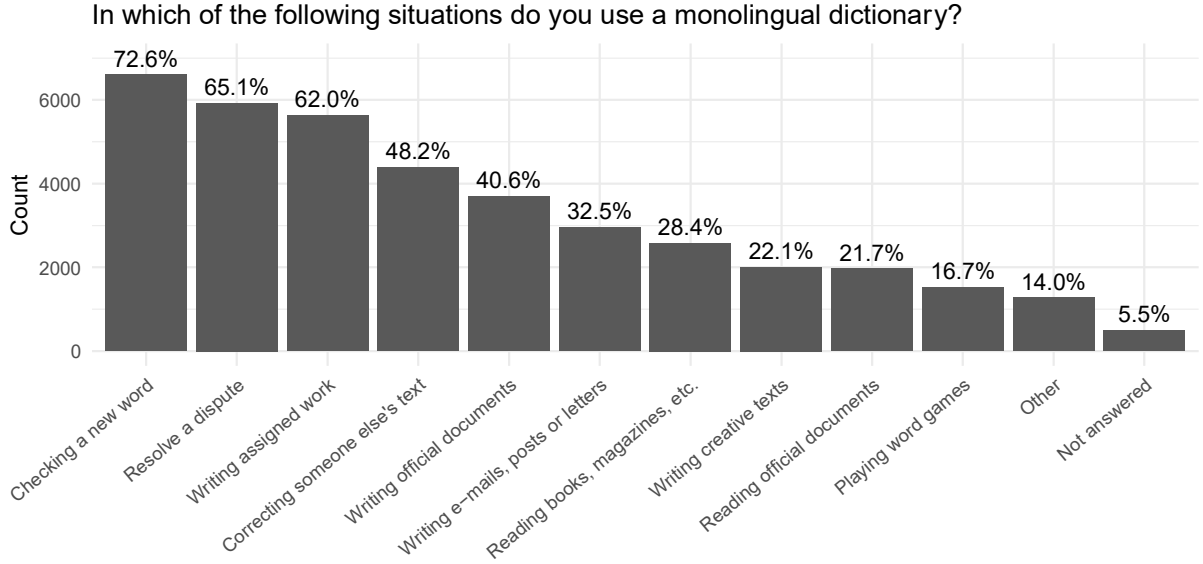

Figure 4: Answer distribution for Question 6 ('In which of the following situations did or do you use a monolingual English dictionary?'). 501 participants were not presented with this

question because they never used a monolingual dictionary. Participants were allowed to select more than one option. Hence, percentages exceed 100%.

Daily situations like checking a new word never heard or read before, as well as more official or work-related tasks like writing assigned work or official documents, were the most frequently selected situations. More leisure-related activities like playing word games were chosen less frequently, but still considerably often. There are no large effects of age group on these usage situations, apart from such effects as can be explained by the more general context of the participants' personal lives (see Section 5 below). For example, people in the youngest age group (24 and younger) tend to use dictionaries less often for correcting someone else's text, while over-55's do not use dictionaries for assignments much.

4.3 *Dictionaries in everyday life*

Questions 7 through 9 were presented to all participants, as it is not necessary to actually use a monolingual dictionary to answer those questions, thus we deliberately included those participants who claimed not to have ever used them. The questions were included to gain an insight into people's everyday opinions of and attitudes towards monolingual dictionaries. We thus presented a mix of concrete situations and more abstract questions regarding dictionaries. A more abstract question was 'Which of the following characteristics do people associate with the leading monolingual [language] dictionaries in [country]?'. **Figure 5** gives an overview of the results. All participants answered this question, presumably because there were two answers ('None of the above' and 'I don't know') that could be checked if one did not want to give an answer. With the exception of these two cop-out options, the order of the response options was randomized to eliminate potential position effects.

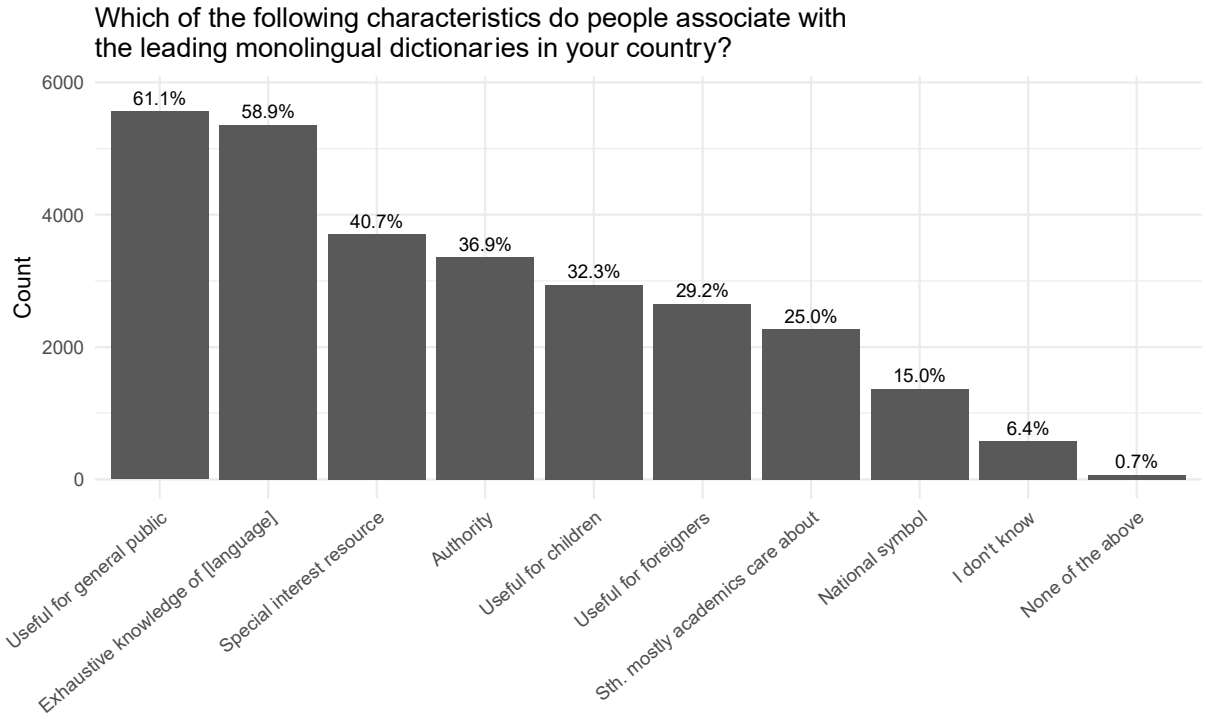

Figure 5: Answer distribution for Question 7. Participants were allowed to choose more than one answer. The full text of the option represented by the second bar from the left was 'Source of exhaustive knowledge of [language]'.

There were small but measurable effects of participants' age for the options 'useful for general public' (r = 0.07), 'exhaustive knowledge of [language]' (r = 0.08), 'authority' (r = 0.09) and 'something mostly academics care about' (r = -0.07), with product-moment correlation coefficients $r > 0$ indicating that older participants were more likely to choose the respective option, and the one negative r indicating that younger participants were more likely to select the option. Since r varies between -1 and 1, all of these effects were weak, but highly significant ($p < 0.0001$), given the large sample size.

Since participants were allowed to select more than one answer, the potential number of different analyses that would be necessary to determine if the country of residence had an effect on the answers is quite large. Basically, one analysis per answer option is necessary and the full presentation of such results would use up a lot of space. Instead, we chose to present, for each option, the three countries that rated the respective option highest and lowest (see **Table 3**). We also include the age-corrected percentage of people from these countries choosing the respective option.[6]

Table 3. Top three and bottom three countries for the different answer options from Question 7. Percentages give the share of participants from the respective country that selected the respective characteristic.

Answer option	Top countries / areas	Bottom countries / areas
Useful for the general public	Basque Country (82.4%) Portugal (79.4%) Denmark (73.8%)	Italy (35.9%) Israel (40.9%) Netherlands (45.3%)
Source of exhaustive knowledge of [language]	Estonia (79.1%) Basque Country (79.0%) Italy (73.2%)	Finland (35.2%) Czech Republic (44.7%) Norway (46.2%)
Resource of special interest	Estonia (76.2%) Romania (60.1%) Netherlands (56.8%)	Basque Country (16.1%) France (20.4%) Croatia (23.1%)
Authority	Netherlands (62.6%) Estonia (60.0%) Finland (54.6%)	Macedonia (11.8%) Georgia (13.4%) Croatia (21.1%)
Useful for children	Greece (55.2%) France (52.3%) Macedonia (47.2%)	Basque Country (12.2%) Finland (20.4%) Poland (20.9%)
Useful for foreigners	France (48.3%) Sweden (47.4%) UK & Ireland (46.2%)	Basque Country (10.3%) Israel (11.8%) Poland (15.8%)
Something mostly academics care about	Czech Rep. (46.5%) Romania (44.3%) Sweden (44.2%)	Basque Country (1.7%) Estonia (5.0%) France (7.8%)
National symbol	Serbia (29.8%) Georgia (28.1%) Slovenia (28.0%)	Portugal (5.5%) Spain (6.4%) Belgium (6.9%)

When we look at the answer option in the last row of **Table 3**, it seems as if people in smaller countries might see their leading monolingual dictionary more as a national symbol than those from bigger countries. However, if we plot the country size and the (again, age-corrected) probability of choosing 'national symbol' (**Figure 6**), we see that this correlation is not as clear as the three highest-percentage countries might suggest. For example, the UK and Ireland as well as Germany show rather high figures for the 'national symbol' option and there are also a range of smaller countries (e.g., Portugal and Belgium) where fewer participants see the leading monolingual dictionary as a national symbol. Spearman's rank correlation coefficient shows a moderate negative relationship between population size and the percentage of people choosing 'national symbol' (ρ = -0.33). However, this effect is not significant (p = 0.11).

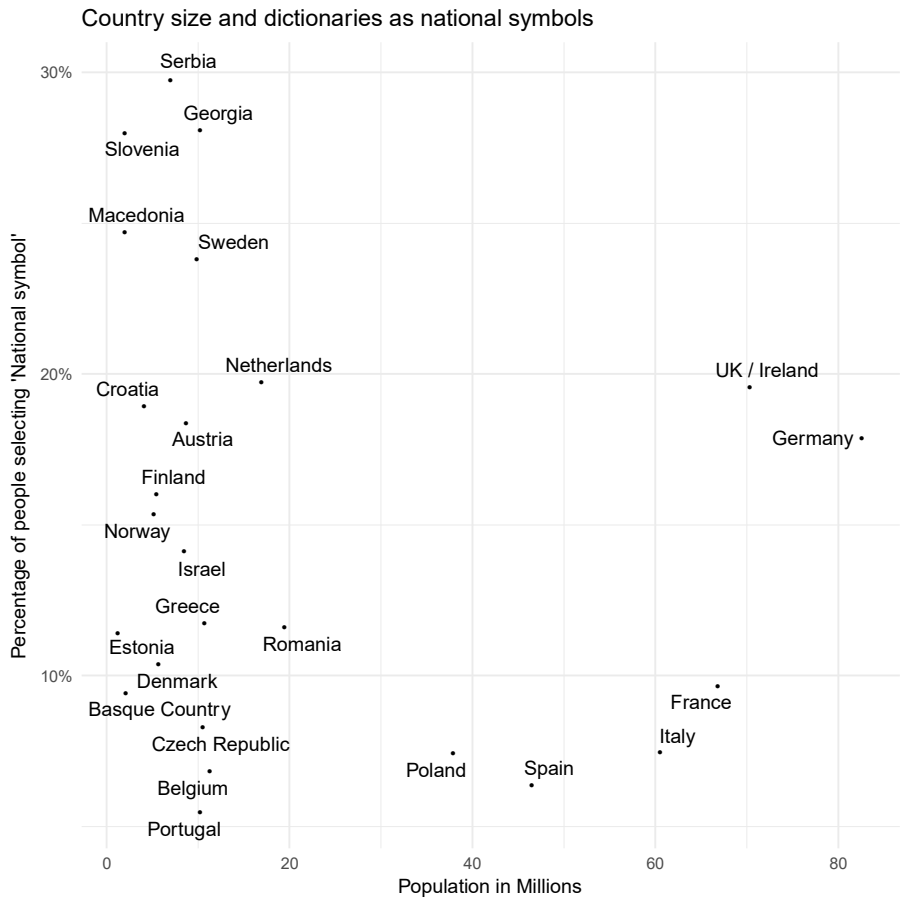

Figure 6: Population of countries in millions (*x*-axis) vs. age-corrected probability of seeing the monolingual dictionary as a national symbol (*y*-axis) for Question 7.

Questions 8 and 9 were aimed at determining attitudes towards dictionaries. We opted for potential real-life situations and presented a range of possible reactions to these situations. Question 8 asked 'Suppose you encounter a word that is not in a monolingual dictionary, then ...'. Participants were again allowed to choose more than one option (except '…none of the above applies'). The distribution of answers is presented in **Figure 7**.

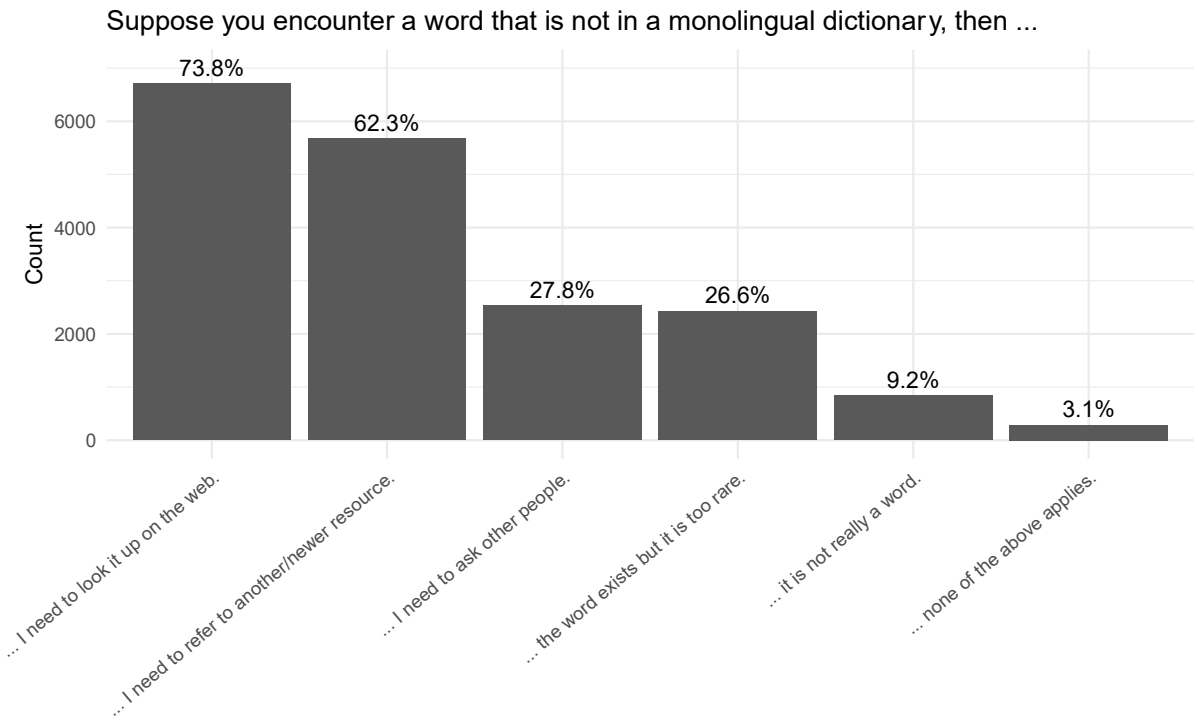

Figure 7: Answer distribution for Question 8. Participants were allowed to choose more than one answer. No participant skipped the question.

It is quite interesting how ubiquitous the Internet is, also in a situation of linguistic doubt. On the other side of the spectrum, it is also quite clear that the vast majority of people are aware of the fact that just because a word is not in a monolingual dictionary, it does not mean that it is not a word. Although 834 participants chose this answer, this only makes for 9.2% of all participants. No substantial age-related effects can be observed for this question (all absolute values of the correlation coefficient are below 0.05, so even significant effects are very small and we will not report them here). The second question tapping into attitudes towards dictionaries was Question 9, asking 'Imagine the following situation: You are cleaning out a relative's house; there you find a well-known monolingual dictionary of [language]. What do you do?' With this question, we hoped to elicit a not too implausible scenario that participants would be able to imagine. The results are summarized in **Figure 8**.

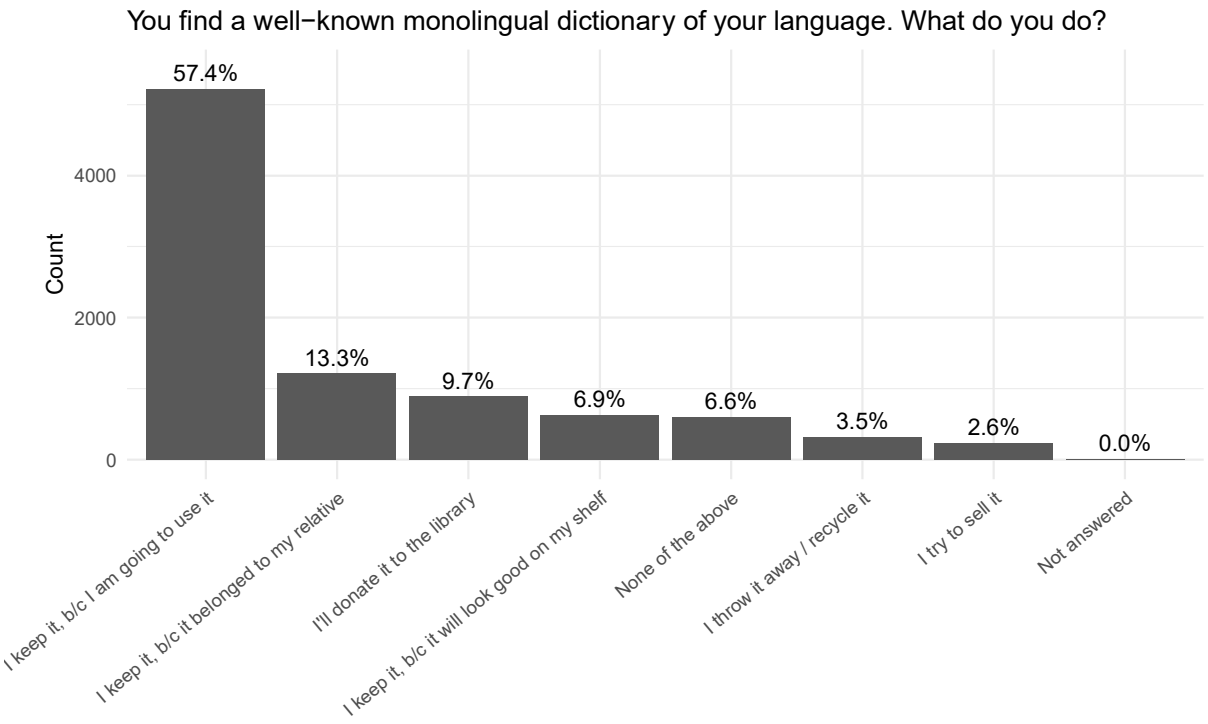

Figure 8: Answer distribution for Question 9. Participants could only choose one answer. Four participants skipped the question (due to rounding, the percentage is reported as 0.0%, the more exact figure being 0.04%).

A fairly large number of people (600, 6.6%) were not satisfied with any of the answer options given and chose 'none of the above' (we had deliberately decided against including a free text option where the participants might specify other courses of action, as this would necessitate manually coding thousands of responses in the many project languages). When summarizing all 'keeping' and 'not keeping' answers, the majority (77.6%) would keep such a dictionary, whereas only 16% would give it away. As can also be seen in **Figure 8**, most participants would keep the dictionary to actually use it and not for reasons of an emotional or aesthetic nature. Age-related effects can be observed for two answer options: older participants were slightly more inclined towards donating it to the library ($r = 0.07$) and younger participants were more likely to keep the dictionary so it would look good on their shelves ($r = -0.14$, both ps < 0.0001).

4.4 *Criteria of a good dictionary*

One of the most important best practices in empirical research has hardly been explored in research into dictionary use: replicating former studies. We attempted to do this to some extent with our question 'What would you find important in a monolingual dictionary of [your language]?' In the study reported in Müller-Spitzer and Koplenig (2014), a very similar question with similar response options had been answered by 684 participants. The data for this study had been collected in 2010, and did not cover the range of countries of the present study. In addition, it focused specifically on online dictionaries. Still, the most highly valued characteristic was *reliability of content*, and so not a feature specific to online or digital dictionaries, but rather a universal quality of reference works. Media-specific features such as adaptive ways of presenting dictionary content, or integrating multimedia features like audio files were ranked and rated as less important, not just in relation to the 'traditional' criteria, but also on an absolute scale. The conclusion of the 2010 study was that the participants wanted an

online dictionary to be a reliable reference work and that medium-specific enrichment or making a dictionary adaptive to different types of user situations was not as important as some lexicographers might think (cf. Müller-Spitzer and Koplenig 2014: 182-186). In the present study, we wanted to see if the situation has changed over the last seven years. The graphs in **Figure 9** directly compare the results from the two studies.

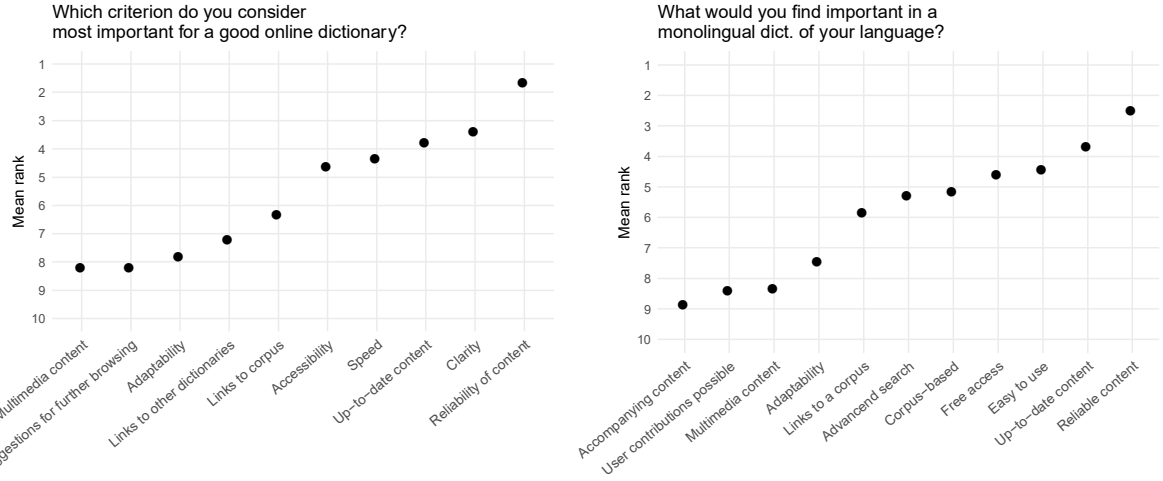

Figure 9: Mean ranks from the Müller-Spitzer and Koplenig (2014) study (left) and mean ranks from the present survey (right). The data presented in the left panel was collected in 2010, while the data in the right panel comes from 2017.

The patterns look very similar, despite the fact that our sample in the present study differs from the 2010 study in several ways. Also, we have to be aware of the fact that some of the criteria are not identical and the number of criteria does not match (2010 study: 10 criteria, 2017 study: 11 criteria). However, there are several striking similarities: reliability of content is still the most important feature, followed by up-to-date content ('clarity' has not been used in the 2017 study). The options to allow user contributions or to use multimedia features are still ranked low on importance. In both studies, linking to corpus data is seen to be of moderate importance (ranked 6th in 2010, 7th in 2017). From these results, a dominant view emerges of the qualities expected of a good dictionary that is rather simple: it should be reliable, up-to-date, easy-to-use, and freely accessible, with other considerations not nearly as important in general. It is also a view that is highly consistent across a range of ages and professional backgrounds. While it holds for most countries, there are departures. Reliability of content appears to be the most robust feature: there is only one country where it is not the top consideration (and, in fact, not even one of top three): Macedonia. Free access is promoted to the top three in a number of countries: Croatia, Estonia, Georgia, Macedonia, Slovenia, and Spain. Linking to corpus is in the top three in Finland, Italy, and Macedonia.

5 General discussion

At the outset, it is encouraging to note that the majority of our participants report using their dictionaries on a daily basis (Question 1), though the issue of sample self-selection needs to be borne in mind when interpreting these results: people who have no interest in dictionaries and do not use them are probably less likely to agree to participate in a survey about dictionaries.

As set out in Section 2, the countries investigated exhibit both similarities and differences in their lexicographic landscapes. It is not surprising that these similarities and differences then translate into similarities and differences in attitudes towards dictionaries and user habits, as revealed by survey responses. As a simple illustration, Serbia has no online monolingual

dictionary available, and so the dominant response as to the medium of choice for dictionary access is print, unlike in our sample overall, where the computer clearly dominates. Computer-based dictionaries are what participants apparently use most (Question 1), and this is also what they prefer (Question 2). A somewhat unexpected outcome is the rather low incidence of dictionary use via the smartphone. To some extent, this may be a consequence of the relatively limited availability of monolingual dictionaries on this platform. On the other hand, given the current tendency towards responsive web design, most online content should be accessible to smartphone users. This, combined with the very low preference for smartphone-based dictionaries (Question 3), suggests that another explanation should be sought, other than the limited availability. It may be that the display size of a mobile phone is found inadequate for efficient dictionary consultation, or perhaps dictionary use often occurs in the context of activities that are computer-based anyway, and so it would be convenient to keep using the same device for dictionary use. Clearly, the matter deserves further study, as discovering the reasons behind scant use of smartphone-based dictionaries may point the way to making them more attractive to users.

It is hardly a surprise that the younger generation tend to use the modern digital formats, while the older participants in our survey still like their paper dictionaries. The difference, however, is not as big as some might expect, except in the use of smartphones by the younger participants, a finding which is also intuitively correct.

Regarding the types of situations in which dictionaries are used, looking up an unknown word and writing come up most frequently. This is quite consistent with previous findings (e.g. Müller-Spitzer 2013), and the age-related effects identified follow naturally from what we know about typical activities in which people in various age groups would normally engage. At one end of the spectrum, younger dictionary users rarely use dictionaries for correcting someone else's text: that is because younger people are rarely found in the role of arbiters of style. At the other end, participants in the oldest age bracket (55 and older) tend to use dictionaries for written assignments less often than participants in other age groups, presumably because people in this age group are generally less likely to be in a situation where they receive assignments.

Our survey suggests that the general image of the dictionary, including its defining characteristics (Questions 7 to 9), has remained largely unaffected by the digital revolution. Still, when interpreting this finding, we need to consider that in quite a few countries (e.g. Georgia and Romania, both with substantial share in our participant pool), traditional print dictionaries still dominate. Even where electronic monolingual dictionaries are available, they are not necessarily among those dictionaries that readily adopt modern features such as corpus integration or user involvement; and when these features are available, users need not necessarily be aware of them, less still care about them. In view of this, it is actually quite surprising that corpus integration enjoys a relatively high rank of importance (and in three countries — Finland, Italy, and Macedonia — it makes it to the top three.

People generally view dictionaries as repositories of knowledge for general use, and only marginally as national symbols. Differences in this last regard do not seem to depend in any clear pattern on either country size or geo-political situation that might imply a need to re-assert national identity, for which purpose a dictionary of the mother tongue might be a useful instrument. It does appear, however, that the view of the dictionary as a national symbol tends to be alive in several countries in which the print format still dominates.

When faced with a lexical problem that does not find immediate resolution in a dictionary, people generally go to the web for help. In fact, studies of log-files indicate that a general web search engine is increasingly the first port of call. The tendency is not new: as early as 2010, half of all the visits for a leading monolingual dictionary of Danish came from Google

(Lorentzen and Theilgaard 2012: 654); yet it is a tendency which is getting stronger. This gravitation towards the digital format and the web is something that dictionary-makers need to take on board; and they need to get on board by designing state-of-the-art, born-digital dictionaries that suit the habits and expectations of 21st-century users. We also have to keep in mind that looking up an unknown word on the web could also lead to an online-dictionary lookup situation: It might well be the case that an online dictionary is among the first few search engine results. In this case, the responses to Question 8 ranked first ('I need to look it up on the web') and second ('I need to refer to another/newer resource') would actually mean the same thing.

If we adopt a bird's-eye view on all the questions, the participants seem to view monolingual dictionaries predominantly as *tools for solving language-related problems*: The dominant use is to look up new words. The dictionaries are seen as useful for the general public and as a source of exhaustive knowledge about a language. Should the dictionary fail to solve a problem in a specific case, alternative resources (the web and/or other resources) are consulted. Most survey participants would keep a monolingual dictionary if they found one, mostly so they could actually use it, rather than for its sentimental value. Responses to the ranking question reveal the qualities that the participants would welcome in a monolingual lexical tool: it should be reliable, up-to-date, and easy to use.

Other possible uses and characteristics of monolingual dictionaries, such as 'recreational' applications (e.g., for playing word games) and symbolic, sentimental, or aesthetic aspects of dictionary use, are indeed present, but they are clearly overshadowed by the utilitarian aspect: the monolingual dictionary is first and foremost seen as a tool.

6 Conclusion

In the foregoing, we have presented the results of the general part of by far the largest-scale survey of dictionary users, with a focus on monolingual dictionaries. The analyses we have presented have mostly been on a question-by-question basis, i.e. we have not explored the full range of possible cross-combinations of variables. With as many as 26 countries in the dataset, plus a number of meta-variables, it is not feasible to present every possible angle in a static research report. However, one potential line of future research would be to find combinations of variables that identify groups (clusters) of users with their own specific needs and attitudes towards monolingual dictionaries. One could think of such a procedure as inducing groups from the responses the participants gave. Other kinds of groups, namely those defined by meta-variables, can already be explored: to enable the readers and the lexicographic community to explore their own specific comparisons, we have built an interactive data explorer (optionally with grouping by meta-variables) for the general survey data, available online at https://owid.shinyapps.io/ESDexplorer (Wolfer et al. 2018).

One general conclusion that is worth re-iterating before closing is that, if dictionaries as lexically-oriented reference tools are to compete with general search engines, then they should move away from print and superficial digitization, and embrace modern technologies by offering truly digital tools for the benefit of the present and future users.

Notes

[1] http://www.elexicography.eu/
[2] UK and Ireland used separate instances of the survey, but were combined into for the purpose of this analysis, thus there were 25 levels of the variable country.
[3] http://www.elexicography.eu/
[4] In order to obtain comparable overviews, basic guidelines with topics and questions that should be addressed were provided.
[5] One possible solution to account for unequal distribution across countries would be to weight the answers of the participants according to the population size of the country. In such a scenario, one would weight all responses from under-represented countries (under-represented in the sense of the number of participants relative to population size) using a value greater than 1, and responses from over-represented countries with a value between 0 and 1. However, in the case of the present study, this would lead to very large weights for large countries (e.g., answers from France would be weighted by a factor of 4.94) and very small weights for smaller countries with comparably high participation rates (e.g., Estonia with 467 participants would be weighted by a factor of 0.05). In other words, figures from France would be weighted over 80 times more than those from Estonia. In the light of these figures, we decided not to weight the answers at all and treat each participant equally. As a result, each participant comes in with the same weight, but some countries might be over-represented.
[6] Age correction was done with a logistic regression model for each of the answer options, with age as a covariate.

References

Atkins, B. T. S. (ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*. (Lexicographica Series Maior 88.). Tübingen: Niemeyer.

Atkins, B. T. S. and K. Varantola. 1997. 'Monitoring Dictionary Use.' *International Journal of Lexicography* 10.1: 1–45.

Béjoint, H. 1989. 'The Teaching of Dictionary Use: Present State and Future Tasks' In Hausmann, F. J., O. Reichmann, H. E. Wiegand and L. Zgusta (eds), *Wörterbücher/Dictionaries/Dictionnaires. An International Encyclopedia of Lexicography, Vol.1*. Berlin: Walter de Gruyter, 208–215.

Benbow, T., P. Carrington, G. Johanessen, F. W. Tompa and E. Weiner. 1990. 'Report on the New Oxford English Dictionary User Survey.' *International Journal of Lexicography* 3.3: 155–203.

Boonmoh, A. and H. Nesi. 2008. 'A Survey of Dictionary Use by Thai University Staff and Students, with Special Reference to Pocket Electronic Dictionaries.' *Horizontes de Lingüística Aplicada* 6.2: 79–90.

Dziemianko, A. 2011. 'Does Dictionary Form Really Matter?' In Akasu, K. and S. Uchida (eds), *Asialex2011 Proceedings. Lexicography: Theoretical and Practical Perspectives*. Kyoto: Asian Association for Lexicography, 92–101.

Dziemianko, A. 2012. 'On the Use(fulness) of Paper and Electronic Dictionaries' In Granger, S. and M. Paquot (eds), *Electronic Lexicography*. Oxford: Oxford University Press, 319–341.

Dziemianko, A. 2016. 'An Insight into the Visual Presentation of Signposts in English Learners' Dictionaries Online.' *International Journal of Lexicography* 29.4: 490-524.

Frankenberg-Garcia, A. 2011. 'Beyond L1-L2 Equivalents: Where Do Users of English as a Foreign Language Turn for Help?' *International Journal of Lexicography* 24.1: 97–123.

Jopling, F. 2003. *Towards a New Methodology for Research on Electronic and Paper-Based Dictionary Use: A Pilot Study of Look-up Patterns on Screen and in Print*. M.A. Thesis, University of Warwick.

Key Results and Cross-Language Comparison. 2012. In: Rehm, G. and H. Uszkoreit (eds), *META-NET White Paper Series*: Springer. http://www.meta-net.eu/whitepapers/overview

Klosa, A., A. Koplenig and A. Töpel. 2014. 'Benutzerwünsche und -Meinungen zu dem monolingualem deutschen Onlinewörterbuch *elexiko*' In Müller-Spitzer, C. (ed.), *Using Online Dictionaries*. (Lexicographica Series Maior 145.). Berlin: Walter de Gruyter, 281–384.

Lew, R. 2002. 'Questionnaires in Dictionary Use Research: A Reexamination' In Braasch, A. and C. Povlsen (eds), *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 12-17, 2002, Vol.1*. Copenhagen: Center for Sprogteknologi, Copenhagen University, 267–271.

Lew, R. 2011. 'Studies in Dictionary Use: Recent Developments.' *International Journal of Lexicography* 24.1: 1–4.

Lew, R. 2015a. 'Opportunities and Limitations of User Studies.' *OPAL - Online publizierte Arbeiten zur Linguistik* 2015.2: 6–16.

Lew, R. 2015b. 'Research into the Use of Online Dictionaries.' *International Journal of Lexicography* 28.2: 232–253.

Lorentzen, H. and L. Theilgaard. 2012. 'Online Dictionaries – How Do Users Find Them and What Do They Do Once They Have?' In Fjeld, R. V. and J. M. Torjusen (eds), *Proceedings of the 15th EURALEX International Congress*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 654–660.

Müller-Spitzer, C. 2013. 'Contexts of Dictionary Use' In Kosem, I. (ed.), *Electronic Lexicography in the 21st Century: Thinking Outside the Paper. Proceedings of the eLex 2013 Conference, 17-19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 1–15.

Müller-Spitzer, C. (ed.). 2014. *Using Online Dictionaries*. (Lexicographica Series Maior 145.). Berlin: Walter de Gruyter.

Müller-Spitzer, C. and A. Koplenig. 2014. 'Online Dictionaries: Expectations and Demands' In Müller-Spitzer, C. (ed.), *Using Online Dictionaries*. (Lexicographica Series Maior 145.). Berlin: Walter de Gruyter, 143–188.

Müller-Spitzer, C., A. Koplenig and S. Wolfer. 2018. 'Dictionary Usage Research in the Internet Era' In Fuertes-Olivera, P. A. (ed.), *The Routledge Handbook of Lexicography*. London: Routledge, 715-734.

Müller-Spitzer, C., S. Wolfer and A. Koplenig. 2015. 'Observing Online Dictionary Users: Studies Using Wiktionary Log Files.' *International Journal of Lexicography* 28.1: 1–26.

Nesi, H. 2000. *The Use and Abuse of EFL Dictionaries*. (Lexicographica Series Maior 98). Tübingen: Niemeyer.

Nesi, H. and K. H. Tan. 2011. 'The Effect of Menus and Signposting on the Speed and Accuracy of Sense Selection.' *International Journal of Lexicography* 24.1: 79–96.

Rundell, M. 2015. 'From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions.' *Lexikos* 25: 301-322.

Tono, Y. 2000. 'On the Effects of Different Types of Electronic Dictionary Interfaces on L2 Learners' Reference Behaviour in Productive/Receptive Tasks' In Heid, U., S. Evert, E. Lehmann and C. Rohrer (eds), *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany*. Stuttgart: Institut for maschinelle Sprachverarbeitung, Universität Stuttgart, 855–861.

Tono, Y. 2001. *Research on Dictionary Use in the Context of Foreign Language Learning: Focus on Reading Comprehension*. (Lexicographica Series Maior 106). Tübingen: Niemeyer.

Töpel, A. 2014. 'Review of Research into the Use of Electronic Dictionaries' In Müller-Spitzer, C. (ed.), *Using Online Dictionaries*. (Lexicographica Series Maior 145.). Berlin: Walter de Gruyter, 13–54.

Welker, H. A. 2010. *Dictionary Use: A General Survey of Empirical Studies*. Brasília: Author's Edition. Accessed on 20 November 2010.

Wolfer, S., I. Kosem, R. Lew, C. Müller-Spitzer and M. Ribeiro Silveira. 2018. 'Web-Based Exploration of Results from a Large European Survey on Dictionary Use and Culture: ESDexplorer.' *Lexikos* 28.