Analysing Translators’ Language Problems (and Solutions)
Through User-generated Content

Jaka Čibej, Vojko Gorjanc, Damjan Popič
Department of Translation, Faculty of Arts, University of Ljubljana
e-mail: jaka.cibej@ff.uni-lj.si, vojko.gorjanc@ff.uni-lj.si, damjan.popic@ff.uni-lj.si

Abstract
This paper focuses on dictionary use among translators and presents the results of a pilot study into translators’ use of language resources when solving language problems. The paper first provides an overview of related work in this field and continues by presenting the results of a pilot study into translators’ use of language resources when solving language problems. By analysing a number of discussions in Prevajaleci, na pomoč!, a dedicated, self-managed Facebook group for Slovene translators aimed at solving translation problems, a taxonomy of typical language problem scenarios is developed. Through an analysis of the problems encountered by translators and their suggested solutions, two aspects are established: the areas in which problems occur and the ways the solutions are reached. This analysis is followed by the final segment, which discusses the suitability of the proposed method and the degree to which this approach yields results that are of use for the compilation of monolingual dictionaries, as well as for lexicographical user research.

Keywords: translators; social media; language resources; monolingual dictionary; dictionary users

1 Introduction
Depending on their professional, linguistic and cultural background, different groups of dictionary users expect to obtain different information from dictionaries. In this paper, we focus on translators. Until now, studies of dictionary use among translators have predominantly focused on using dictionaries in translator training, or training translators in the use of dictionaries (Roberts 1992, Sanchez Ramos 2005, Hirci 2013), as well as using dictionaries in professional translator work (Nuccorini 1992, Roberts 1997, Atkins & Varantola 1998, Varantola 2002, Sanchez Ramos 2005). The methodology of initial dictionary user research often consisted of questionnaires, while studies of the translation process employed the TAP (think-aloud protocol) method, which requires the translator to verbalise their dilemmas and decisions aloud while translating. Today, the TAP method has been partially supplanted by more advanced approaches such as screen capture or keylogging with specialised software (e.g. Translog1), or even eye-tracking (Hirci 2009; see also Tono 2011). At least in the case of translators, studies on their use of language resources are no longer isolated, but often form a part of broader interdisciplinary studies of the translation process (Paulsen Christensen 2011).

With language students, dictionary user research often involved the use of dictionaries in second language acquisition, particularly English as a foreign language (Béjoint 1981, Mackintosh 1998, Humblé 2001). These studies were followed by further research into dictionary use in translator training (Roberts 1992, Sanchez Ramos 2005, Hirci 2013), mostly dealing with different aspects of

dictionary use in language acquisition or translation – predominantly by analysing user experience, which led to plans and suggestions for a more systematic inclusion of dictionary information in the learning process.

In this paper, we discuss translators’ user needs based on an analysis of user-generated posts published in the Facebook group *Prevajalci, na pomoč!* (‘Translators, help!’), which enables Slovene translators to discuss language dilemmas they encounter in their work. We focus only on posts pertaining either exclusively to the use of Slovene (the native language of the majority of translators in the group) or bilingual questions with Slovene as the target language.

### 2 Motivation

In Slovenia, lexicographical user studies are scarce, having emerged only recently (Arhar Holdt et al. 2016a, Arhar Holdt et al. 2016b, Mikolič 2015, Rozman et al. 2015). A number of them (Hirci 2013, Gorjanc 2014, Čibej et al. 2015) also focus on translators and the language resources they use (and need). While bilingual dictionaries are among the most obvious, translators also consult monolingual language resources (e.g. general monolingual dictionaries and terminological resources). The main motivation for the pilot study presented in this paper is the preparation of a new digital monolingual dictionary of Slovene (Gorjanc et al. 2015), the aim of which is to meet the needs of diverse groups of dictionary users, including (professional) translators (Čibej et al. 2015). As Slovene translators are under-resourced not only in terms of bilingual dictionaries, but language resources in general, we intend to show that a number of Slovene translators’ language needs should be taken into account when designing a monolingual dictionary in order to create a versatile language resource that is useful even in a community dealing with more specialised language dilemmas.

### 3 Observation of Translator’s Needs Through Facebook Posts

The Slovene translator community has shown itself as well organised in social media through the Facebook group *Prevajalci, na pomoč!* Many of the group’s members participate daily in discussions of actual translation dilemmas they encounter in their work. We thus hypothesise that a systematic analysis of user posts could reveal the manner in which translators address individual translation dilemmas, as well as to what degree specific language resources could contribute to their resolution.

#### 3.1 Methodology

First, we manually collected language-related questions posted in the Facebook group *Prevajalci, na pomoč!* in the period between November 2014 and January 2015. Our goal was to collect 100 posts relevant to our study. From the initial collection of 223 most recent posts, we eliminated all irrelevant material (e.g. job offers, out-of-topic posts, and questions with a foreign target language). In the final selection containing 100 relevant posts, several included more than one question. We divided the posts into questions and ended up with a dataset of 106 questions in total. The questions were then classified into categories representing different scenarios, designed bottom-up (based on the material). We present the categories and provide examples below.
Scenario 1: *I have no clue and I need a translation.* → The translator knows nothing about a foreign language expression and requires help understanding it and finding a Slovene equivalent, as shown in example (1).²

(1) Untreated redwood – what kind of wood is that (if we even have it) and how do we translate it?

Scenario 2: *I know the concept, but not the term/expression.* → The translator is already familiar with the concept in the foreign language in question, but requires help verbalising it in Slovene. In this scenario, it is common that the translator already provides a Slovene equivalent, most often as a description, as shown in example (2).³


Scenario 3: *I already have a solution to my problem. Is it accurate?* → The translator has found a Slovene equivalent, but doubts its correctness and requires confirmation, as shown in example (3).⁴

(3) heritage sites – dediščinske točke (??) site of interest (museums, galleries, castles, caves, etc.) – točke interesa (??)

Scenario 4: *I’ll choose the right expression myself. I need ideas.* → The translator has already decided to determine an adequate expression by themselves. In such cases, the Slovene expression is either not yet in use or not described in language resources. The aim of the translator could also be to express creativity (esp. when translating e.g. literature or films). An example is shown in (4).⁵

(4) How do you translate “performance” in the context of make-up/cosmetics? In my case, there’s the expression “…impeccable performance of make-up products…”. Obstojnost, učinkovitost? What should I use in Slovene in this case?

Scenario 5: *I have two or more options. Which one is better?* → The translator has collected a number of solutions but is not certain which one is adequate (in terms of meaning, orthography, etc.), as shown in (5).⁶

(5) Is the Slovene version of the word ‘skeeball’ already in use, e.g. skibol (like bejzbol)? I highly doubt it, but I thought I’d ask. Thank you.

---

² Example (1) in Slovene: Untreated redwood, kateri les je to pri nas (če sploh je pri nas) oz. kako se prevaja?
⁴ Example (3) in Slovene: heritage sites – dediščinske točke (??) site of interest (muzeji, galerije, gradovi, jame ipd.) – točke interesa (??)
⁵ Example (4) in Slovene: kako prevajate “performance” v kontekstu ličil/kozmetike? v mojem primeru govori o “…impeccable performance of make-up products…”. Obstojnost, učinkovitost? Kaj bi se uporabilo v slo. na tem mestu?
⁶ Example (5) in Slovene: Se ‘skeeball’ slučajno že sloveni kot recimo skibol (kot bejzbol)? Sicer močno dvomim, a vseeno vprašam. Hvala.
Scenario 6: My solution is not adequate. I want an alternative. The translator has found a solution they know to be inadequate for some reason. They seek an alternative. An example is shown in (6).

(6) Just out of curiosity: does anyone know if there’s a better translation for “think tank” than the uninspired, albeit commonly used ‘možganski trust’?

Scenario 7: What does this expression mean? The translator has encountered a Slovene expression and requires a definition to understand it (see example (7)).

(7) Does anyone know what “tramak” means? The context: se omejnost skladiščnih površin rešuje s tramaki blaga.

Scenario 8: What do you think of the solution provided by my language editor? The translator does not agree with the solution provided by the language editor and seeks either an explanation for the editor’s decision or a confirmation that the editor’s correction is indeed illogical (see example (8)).

(8) LOL, my language editor killed me when she said that the Slovene Manual of Orthography recommends the use of ‘odtisoček’. What do you think?

Scenario 9: I need an expression in standard Slovene. The translator has found a solution to their problem in the form of a non-standard (e.g. colloquial) expression, which cannot be used in the translation because of limitations concerning register and/or style (see example (9)).

(9) Does anyone maybe know the standard equivalent of ‘prskalica’? So far, I’ve only got ‘carobna svecka’, but it seems dubious.

Scenario 10: I want to know the etymology of this expression. The translator wants to know the origin of a certain expression (see example (10)).

(10) Does anyone have an etymological dictionary at hand’s reach or happen to know the origin of the word KMET?

The categorisation was followed by a thorough qualitative analysis of the collected questions (and the replies provided by group members). We present the highlights of the results in the following sections.

---

7 Example (6) in Slovene: Takole iz firbca me zanima, ali kdo ve, če se je za “think tank” že nasla boljša rešitev od precej neposrečenega, čeprav dokaj uveljavljenega “možganskega trusta”? 
8 Example (7) in Slovene: a kdo ve, kaj pomeni »tramak«? Kontekst: se omejnost skladiščnih površin rešuje s tramaki blaga.
9 Example (8) in Slovene: LOL, lektorca me je ubila, ko je povedala, da SP predlaga odtisoček. Kaj pa menite vi?
10 Example (9) in Slovene: mogoce kdo ve, kako po slovensko recemo prskalici? Zaenkrat sem nasla samo “carobna svecka”, ma me ne preprica.
11 Example (10) in Slovene: ima morda kdo pri roki etimološki slovar ali informacijo o izvoru besede KMET?
3.2 Results

Table 1 shows the scenarios used in question categorisation and their percentages.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1: I have no clue and I need a translation.</td>
<td>20</td>
<td>19%</td>
</tr>
<tr>
<td>Scenario 2: I know the concept, but not the term/expression.</td>
<td>33</td>
<td>31%</td>
</tr>
<tr>
<td>Scenario 3: I already have a solution to my problem. Is it accurate?</td>
<td>14</td>
<td>12%</td>
</tr>
<tr>
<td>Scenario 4: I’ll choose the right expression myself. I need ideas.</td>
<td>11</td>
<td>10%</td>
</tr>
<tr>
<td>Scenario 5: I have two or more options. Which one is better?</td>
<td>10</td>
<td>9%</td>
</tr>
<tr>
<td>Scenario 6: My solution is not adequate. I want an alternative.</td>
<td>9</td>
<td>9%</td>
</tr>
<tr>
<td>Scenario 7: What does this (Slovene) expression mean?</td>
<td>4</td>
<td>4%</td>
</tr>
<tr>
<td>Scenario 8: What do you think of the solution provided by my language editor?</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td>Scenario 9: I need an expression in standard Slovene.</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td>Scenario 10: I’m interested in the etymology of this word.</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Total</td>
<td>106</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1: Question categorisation according to scenarios.

The overview of various scenarios and their percentages shows that translators most commonly encounter language problems stemming from expressions they either do not understand or, more often, cannot (adequately) verbalise. In order to design a monolingual dictionary that would contribute to the solution of these problems, the needs indicated by the above scenarios need to be taken into account. Apart from the needs demonstrated by Scenario 1, which cannot be met by a monolingual dictionary of Slovene as they pertain to problems with foreign language expressions, in all the other scenarios, which focus more on the Slovene part of translation, the monolingual dictionary could prove useful, as the questions often involve a comparison of two or more options. As can be deduced from most of the questions in Scenarios 2–9, translators most often have problems not in decoding the original, but rather in verbalising the translation. In the following section, we focus on the quantitative analysis of the questions in which a monolingual language resource could help solve the problem in question.

3.2.1 Potentially Helpful Dictionary Sections

In addition to categorising the material according to scenarios, we also annotated each relevant question with dictionary information that could potentially help the translator solve the language problem at hand. The results are shown in Table 2.

<table>
<thead>
<tr>
<th>Potentially helpful dictionary sections</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminology</td>
<td>35</td>
<td>32%</td>
</tr>
<tr>
<td>Synonyms</td>
<td>27</td>
<td>24%</td>
</tr>
<tr>
<td>Dictionary entry of problematic words</td>
<td>25</td>
<td>22%</td>
</tr>
<tr>
<td>Labels (register, context, etc.)</td>
<td>9</td>
<td>8%</td>
</tr>
<tr>
<td>Word families</td>
<td>7</td>
<td>6%</td>
</tr>
<tr>
<td>Description of meaning</td>
<td>3</td>
<td>3%</td>
</tr>
<tr>
<td>Examples of use</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td>Etymology</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Idioms</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Orthographic rules</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Total</td>
<td>111</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 2: Percentages of dictionary sections perceived as conducive to a solution.
As shown in Table 2, practically all dictionary information is relevant for translators. It is interesting to note that examples of use rarely contribute to a solution, as dilemmas encountered by translators often involve very specific contexts not covered by the examples of use included in the dictionary. In addition to dictionary entries of problematic words (e.g. neologisms often not included in existing dictionaries), two other categories seem particularly significant for translators: terminology and synonymy.

The fact that terminology is among the most commonly searched for categories of information among translators opens a discussion on the quantity of specialised vocabulary to be included in a general monolingual dictionary, a problem also pointed out by Vintar (2015), who distinguishes between several different term categories according to their termhood – the degree to which the expression is specialised, i.e. used only in a certain field. In the case of this pilot study, it remains uncertain whether a general dictionary could indeed help solve the terminological problems faced by translators, as these may in fact only be solved using a specialised terminological language resource. On the other hand, a monolingual general dictionary could certainly help resolve the question of synonymy. Nevertheless, various aspects need to be considered: how should synonymy be presented in the dictionary? What is the relationship between a general dictionary and a thesaurus? And finally, how should this type of information be integrated into one language resource to make it accessible to the user in an intuitive manner?12

3.2.2 Methods of Problem Solving

We also observed the participants involved in solving the language problem, as well as the sources used (if any). The results are presented in Table 3.

<table>
<thead>
<tr>
<th>Methods of problem solving</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translators</td>
<td>96</td>
<td>63%</td>
</tr>
<tr>
<td>Other (language) resources</td>
<td>30</td>
<td>20%</td>
</tr>
<tr>
<td>Browsing the Internet</td>
<td>19</td>
<td>13%</td>
</tr>
<tr>
<td>Experts</td>
<td>6</td>
<td>4%</td>
</tr>
<tr>
<td>Total</td>
<td>151</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 3: Sources used in problem solving.

As the group is meant for discussion among translators, it is logical that in most cases, the translators themselves are the ones that provide a solution to the problem at hand. However, many solutions are based on information found on the Internet in general. As shown in Table 4, translators explicitly reference certain language resources.13

---

12 Many online monolingual dictionaries, especially English ones (e.g. Oxford English Dictionary, Merriam Webster, Collins, Macmillan), offer both a monolingual dictionary and a thesaurus, either explicitly (both are listed among the sources) or implicitly (the user performs a search and the list of hits includes a synonym section).

13 We only included the sources that were directly mentioned or suggested in the discussion. It is likely that the translators also consulted numerous other sources, so the list is not comprehensive.
Websites seem to be the most conspicuous, as translators often refer to them when searching for terminology. As it turns out, the starting point of problem solving is often a web browser (in our case Google). This is typical of the general modus operandi of translators working with online resources: they start their search for a solution not by directly consulting a specialised source, but with a general Internet search.\textsuperscript{14} 

It is interesting to note that despite the relatively large number of questions that can only be resolved with additional contextual information (provided e.g. by corpora), the use of corpora by translators is negligible. This might indicate either that translators do not consult corpora in their work or that they are not familiar with this type of language resource.

4 Conclusion

The goal of our study was to determine whether the analysis of Facebook posts containing verbalised translation problems can indicate the type of information that is relevant for translators and as such should be included in monolingual (dictionary) resources. The method has proven to be successful as it provided a relatively thorough overview of user needs, categories of content-specific problems, and methods of problem solving.

Even though many different types of lexicographically relevant information are used in resolving translation problems, the aspect that warrants the most attention based on our analysis is the inclusion

\textsuperscript{14} This is a general trend, as can be deduced from the results of related studies. As shown by Lorentzen & Theilgaard (2012), 49% of users access the Danish dictionary \textit{Den danske ordbog} through a web browser, either directly (33%) or through third-party websites (18%). When the site was optimised in terms of search-engine indexing, the web browser access rate rose to 84%.
of a) specialised vocabulary, and b) synonymy.
The first question entails the decision on the degree of specialised vocabulary in a monolingual
dictionary and opens a discussion on whether terminological questions can really be solved by any
other source than a specialised dictionary. This warrants a more systematic and focused analysis of
terminological problems faced by translators, especially regarding the level of termhood exhibited by
the problematic expressions.

As for synonymy, it is clear that translators expect a modern language resource to include information
typical of both a traditional dictionary and a thesaurus. The results show that resolving these two
dilemmas in the new dictionary of Slovene would significantly contribute to its usefulness in
resolving translation problems.
The manner in which translators search through sources indicates that the presentation of information
in a dictionary warrants a completely new approach, as the traditional distribution of information
turns out to be inadequate for the digital medium. For instance, a recurring problem of potential
dictionary users in both our study and the study conducted by Arhar Holdt et al. (2016b) has turned
out to be the need to compare multiple options (with translators, this often involves synonyms,
especially when they seek an appropriate expression in terms of style or register). This need strongly
supports a departure from the traditional dictionary interface and calls for the incorporation of a
comparative option that allows for the juxtaposition of two or more dictionary entries (or their parts).
Inspiration for such a function could perhaps be drawn from other digital resources such as web
browsers, which enable users to open websites in multiple tabs and/or windows.

As the pilot study has shown, the translators’ first impulse is to search in the manner shaped by
modern web browsers: they generally tend to search for information by choosing a general browser as
their starting point, gradually moving from a general resource that covers a wide range of possibilities
to a more specific source (websites, dictionaries, terminological databases, etc.).
In addition, it is necessary to devise ways to make sure that the new dictionary will be as up-to-date
with modern language use as possible. In many cases included in our study, the translators
encountered problems dealing with words that have not yet been described in existing language
resources (e.g. neologisms). As the most widely used Slovene language resources are either in printed
form or, at best, in a digital manifestation of their printed form, they are often outdated. A digital
medium allows for quick updates, but the criteria for the inclusion of new words will have to be
determined.

In the paper, we presented a method to analyse the language needs of translators by observing their
verbalised problems through a Facebook group. Although it can only be considered a pilot study (and
even though it focuses only on dilemmas pertaining to Slovene), it provides interesting insight into
the main problematic points that could be alleviated with the help of a well-designed monolingual
dictionary. For a more comprehensive overview of translator needs, further studies are required, with
methodologies ranging from traditional questionnaires and interviews to studies in which the
scenarios described in this pilot study can be tested on target groups in an authentic translation
environment. This will provide more reliable data on (potential) dictionary users and their
expectations with regard to language resources in general as well as the new monolingual dictionary
of Slovene.
5 References


traduction, terminologie, rédaction 1/5, pp. 49–76.


