# Seeking health content online:
# A survey of Internet users' habits and needs

Alessandra Rossetti

*SALIS, Dublin City University (Ireland)*

Sharon O'Brien

*ADAPT / SALIS, Dublin City University (Ireland)*

**Abstract**

This paper describes a small-scale survey conducted among non-native speakers of English living in Ireland. We collected data from 86 respondents by means of an online questionnaire. Our goal was to investigate their health information seeking behaviour; potential comprehension issues with health content; and their adoption of machine translation (MT) systems. We found that: the Internet is widely used for health-related searches; Wikipedia is the most consulted website; and information on illnesses and public health threats is frequently sought. We observed that the language in which online searches are conducted is influenced by respondents' self-reported level of English proficiency, with most limited English proficiency (LEP) Internet users looking for health information in their native languages to facilitate comprehension. We also observed that specialised medical vocabulary might hinder comprehension. Finally, most participants reported adopting MT to translate online health content from English into their native languages, and LEP respondents reported using MT more frequently than proficient respondents. This survey highlights the need to increase the accessibility of online health content for non-native speakers of English (especially LEP users) with a view to reducing their vulnerability. We argue that text simplification might lead to the production of more comprehensible and more machine translatable health-related texts.

*Keywords:* online health information; non-native speakers of English; comprehension; machine translation

## 1. INTRODUCTION

Despite the increasing linguistic diversity that characterises Irish society, and the need to provide accessible online health information for different language groups (Health Service Executive, 2012), there seems to be a lack of studies on the online health information seeking behaviour of non-native speakers of English living in Ireland, as well as on their specific needs.

Previous works have mainly focused on Irish society as a whole, and did not include machine translation (MT) use. From a survey on European citizens' digital

health literacy (European Commission, 2014), it emerged that most Irish respondents are satisfied with online health information and do not have comprehension issues. Gallagher et al. (2008) found that gender, age, education level and employment status have an impact on online health information seeking in Ireland. However, the authors reported the inability to include immigrants or refugees among the limitations of their study.

Our small-scale survey aimed at partially filling this research gap by specifically targeting non-native speakers of English living in Ireland, with a view to gathering data on their habits and communication-related needs (including MT use) when searching for healthcare information online. In Section 2, we will present related literature. We will then describe the methodology (Section 3) and the data analysis (Section 4). In Section 5, we will discuss the main findings and implications of this study.

## 2. RELATED WORK

The Internet is an important source of healthcare information (Tan and Goonawardene, 2017) that can reduce disparities, but only if content is understood (Berland et al., 2001). Research has shown that is not the case for all Internet users.

Comprehension of English medical content might prove challenging, particularly for non-native speakers of English with a low level of health literacy, defined as "people's knowledge, motivation and competences to access, understand, appraise and apply health information" (World Health Organisation 2013, p. 4). Mcinnes and Haglund (2011) showed, for instance, that most online health information would be unreadable for Internet users with low literacy levels. The authors also observed that content on Wikipedia pages was amongst the most difficult to understand.

Furthermore, most online health content is available in English only. This imbalance represents a barrier for around six billion people who do not have English as their first or second language (Adams and Fleck, 2015). Of the 155,000 medical articles available on Wikipedia at the end of 2013, around 19% were in English, while the rest was distributed across 255 natural languages (Heilman and West, 2015). At the European level, Internet users from the Czech Republic, Latvia, Bulgaria, Poland and Finland reported a lack of availability of online health information in their native languages (European Commission, 2014).

Translation initiatives are underway. For instance, the World Health Organisation's (WHO) website is available in six different languages. However, a divide still exists between users who have access to online health information and those who do not have it (Adams and Fleck, 2015) since translating medical texts is a costly and time-consuming activity (Gan, 2012).

Free, web-based MT systems, such as Google Translate or Microsoft Bing Translator, have been considered as an alternative to human translation (Taylor et al., 2015). Kirchhoff et al. (2011) showed that combining MT with human editing results in time savings, while retaining the same quality of human translation. However, raw MT outputs are often characterised by errors or inaccuracies which might have detrimental effects on the well-being of patients (Wołk and Marasek, 2015).

## 3. METHODOLOGY

### 3.1 Design

We developed an online questionnaire on Google Forms. Since we could not translate the survey into the various languages of the potential respondents, only individuals with the ability to comprehend information in brief, ordinary English texts were able to complete the questionnaire.

One question, which dealt with respondents' self-reported level of English proficiency, was aimed at identifying LEP individuals, namely individuals who report speaking English less than very well (Pandya et al., 2011). Vickstrom et al. (2015) provided some evidence of the validity of this question by showing that answers can differentiate between different levels of literacy.

### 3.2 Targeted participants and survey distribution

This survey was circulated online from August 30, 2016 to October 28, 2016. We targeted non-native speakers of English who lived in Ireland and were 18 or over.

We adopted a random sampling method and contacted those organisations and institutions that were likely to provide access to immigrants living in Ireland, such as: Irish-based embassies of countries where English is not the official language; organisations providing services and support to immigrants in Ireland; or schools running adult literacy courses for non-native speakers of English. Overall, questionnaires completed and submitted by 86 eligible respondents were analysed.

## 4. DATA ANALYSIS AND RESULTS

### 4.1 Background characteristics

The respondents had different language backgrounds. The majority were from Latvia and Slovakia (Figure 1). The native languages most frequently reported were Latvian, Slovak, and Russian (Figure 2). Several participants were either bilingual or multilingual.
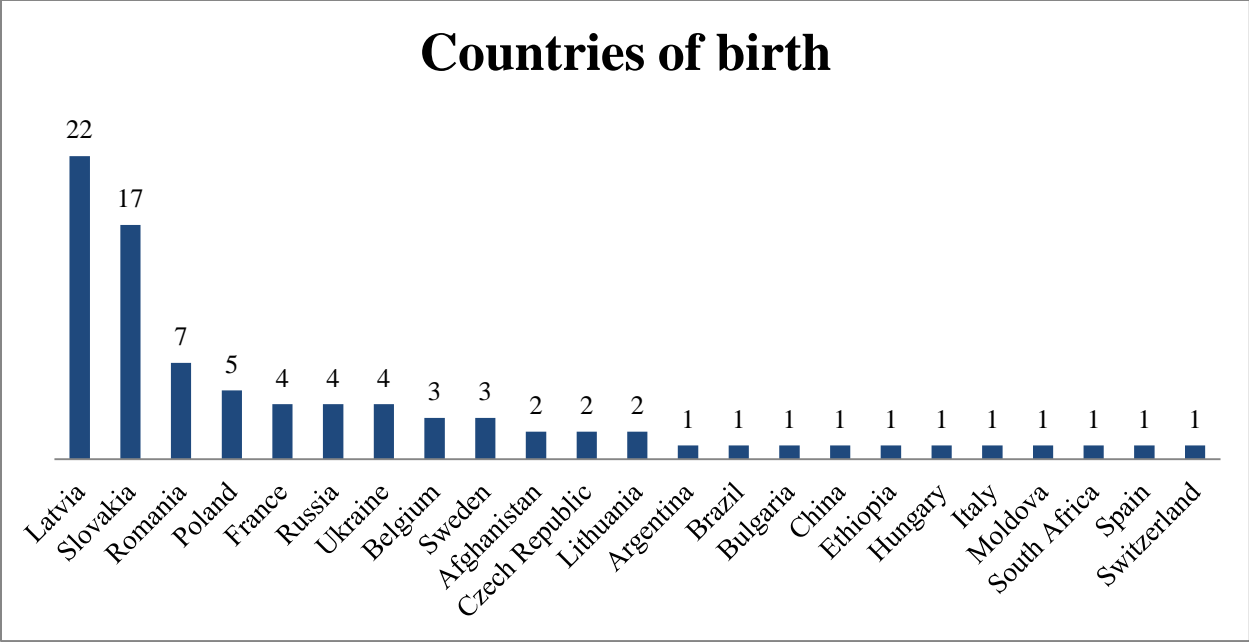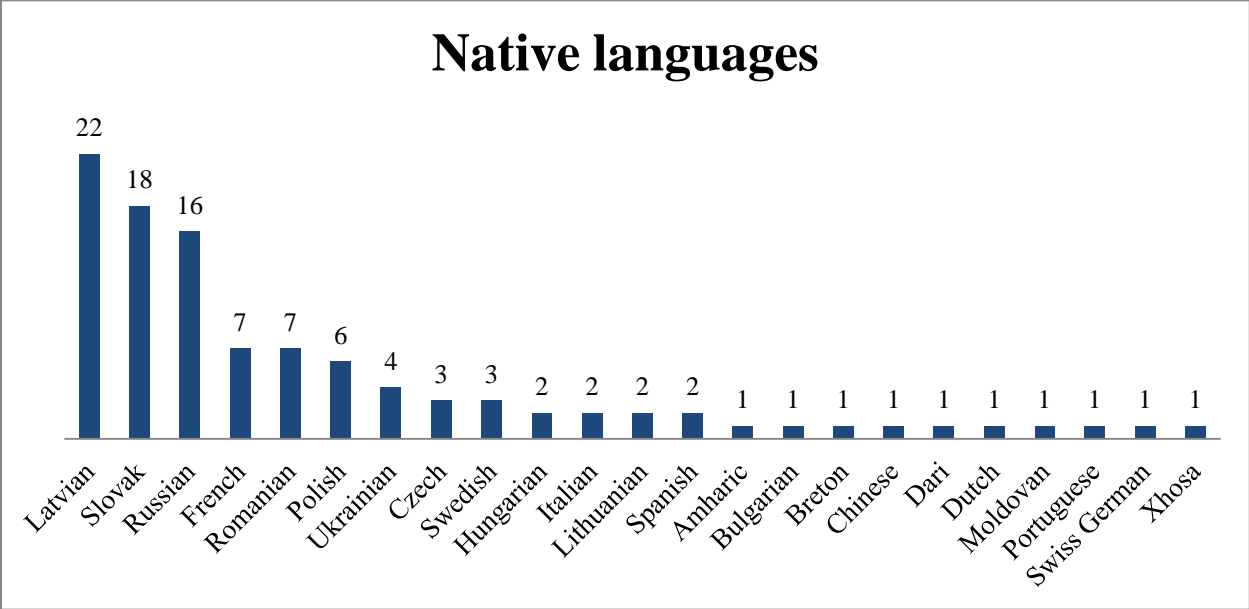
## Countries of birth

Latvia 22, Slovakia 17, Romania 7, Poland 5, France 4, Russia 4, Ukraine 4, Belgium 3, Sweden 3, Afghanistan 2, Czech Republic 2, Lithuania 2, Argentina 1, Brazil 1, Bulgaria 1, China 1, Ethiopia 1, Hungary 1, Italy 1, Moldova 1, South Africa 1, Spain 1, Switzerland 1

Figure 1. Countries of birth

## Native languages

Latvian 22, Slovak 18, Russian 16, French 7, Romanian 7, Polish 6, Ukrainian 4, Czech 3, Swedish 3, Hungarian 2, Italian 2, Lithuanian 2, Spanish 2, Amharic 1, Bulgarian 1, Breton 1, Chinese 1, Dari 1, Dutch 1, Moldovan 1, Portuguese 1, Swiss German 1, Xhosa 1

Figure 2. Native languages

More than half of the 86 respondents were aged 18 to 38 (Figure 3). The majority were women (Figure 4).

## Age range

■ [18-38]　■ [39-59]　■ [60-80]

(2) 2%

28 (33%)

56 (65%)

Figure 3. Age ranges

## Gender

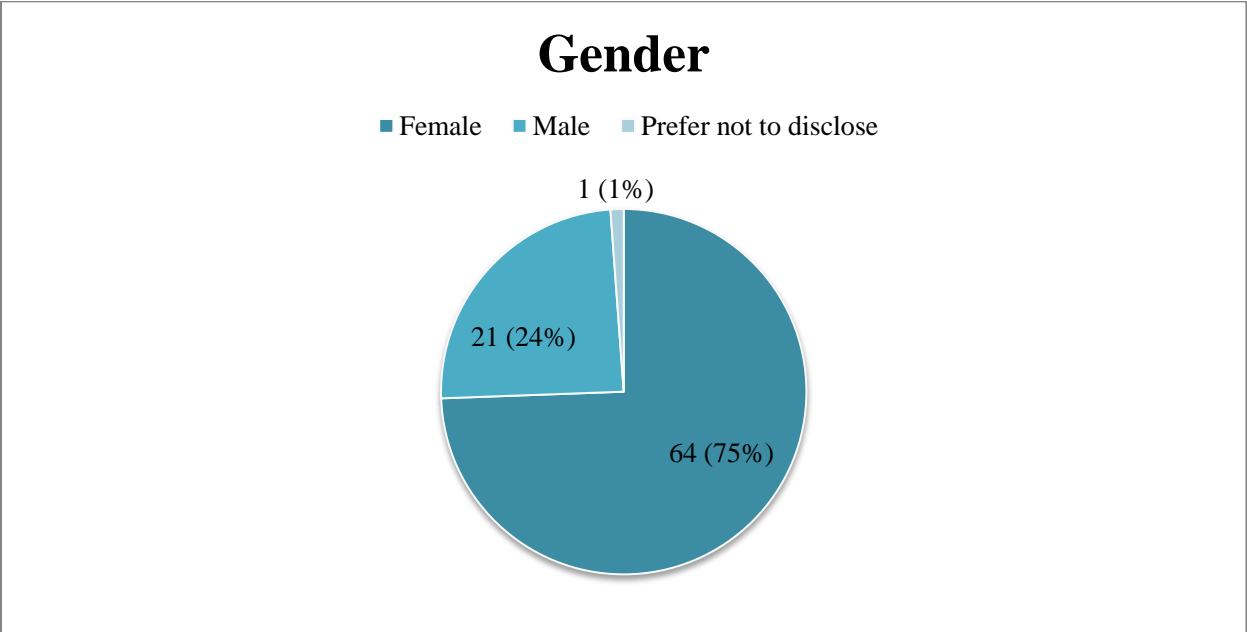■ Female　■ Male　■ Prefer not to disclose

1 (1%)

21 (24%)

64 (75%)
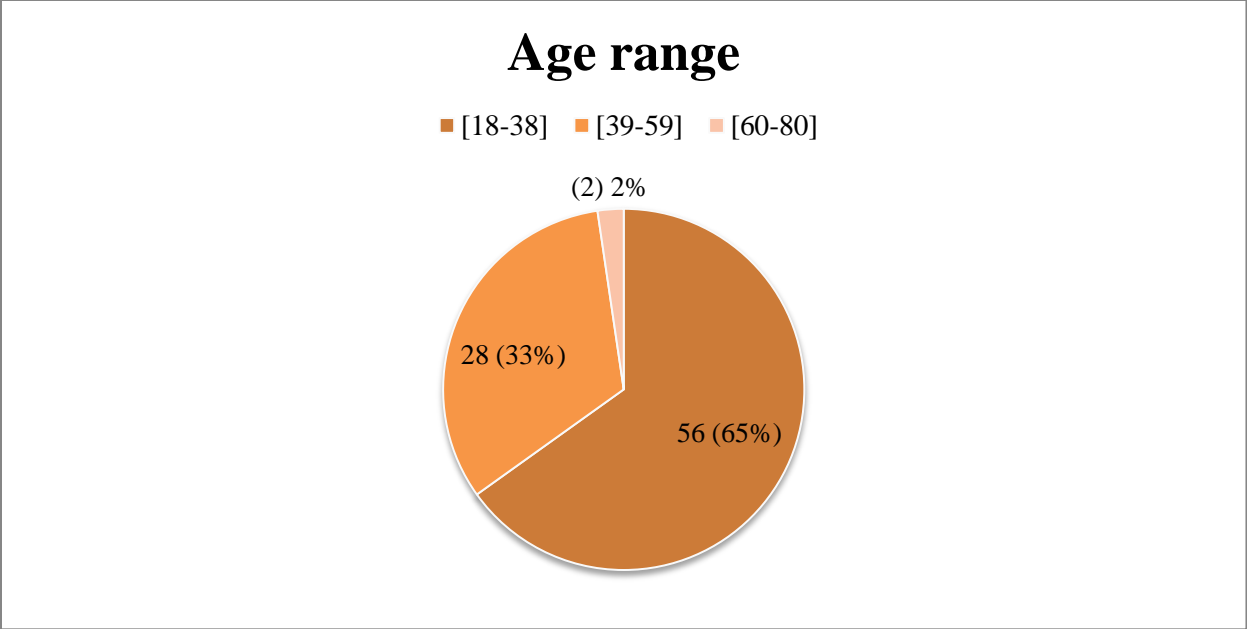
Figure 4. Gender distribution

Respondents had moved to Ireland between 1992 and 2016 (Figure 5). One response on year of arrival was uninterpretable. Regarding self-reported level of English proficiency (Figure 6), more than half of the 86 respondents reported speaking English very well (henceforth *proficient respondents*). 46% of them

reported speaking English less than very well, i.e. either well, not well or not at all (henceforth *LEP respondents*).

**Years of arrival in Ireland**



Figure 5. Years of arrival in Ireland

**Self-reported level of English proficiency**

■ Very well  ■ Well  ■ Not well  ■ Not at all
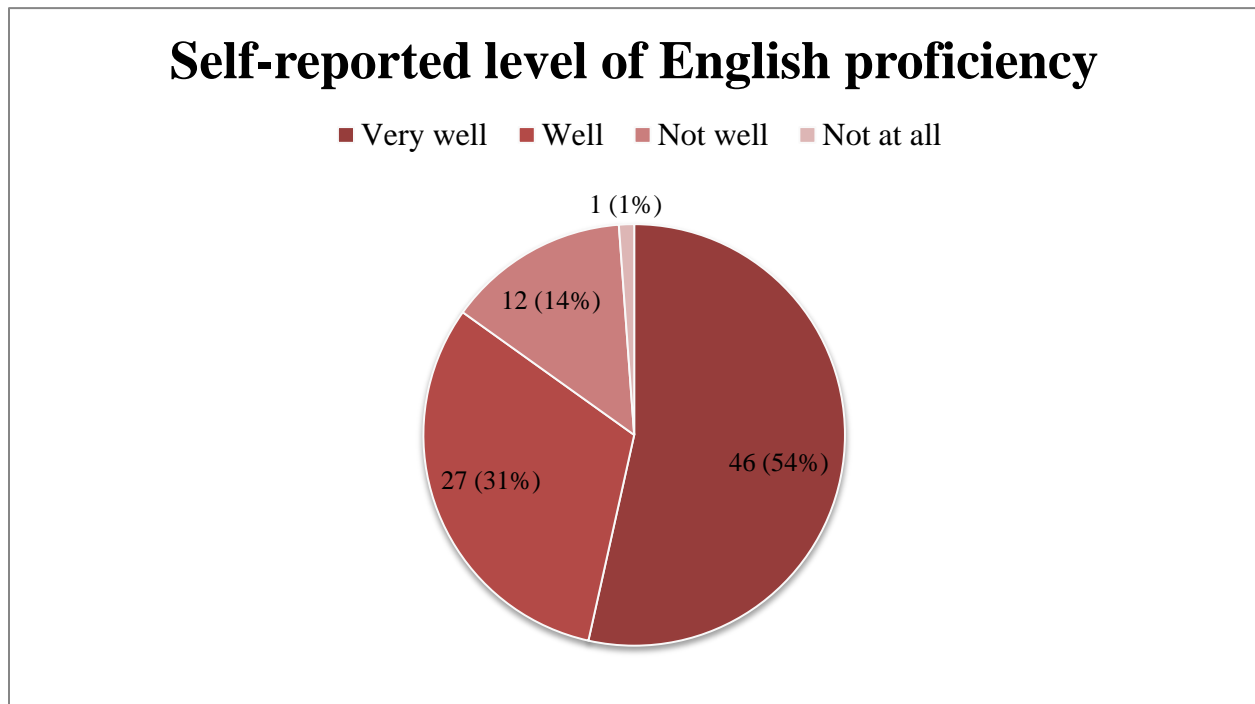


Figure 6. Self-reported level of English proficiency

### 4.2 Internet use

The vast majority of the participants (81) reported using the Internet to seek online health information (Figure 7). The remaining five participants provided a

variety of reasons for not consulting the Internet, including: availability of health practitioners, lack of trust in online health content, and inaccuracies. These five participants were asked to skip the following questions. Therefore, the next answers were collected from a sample of 81 respondents.
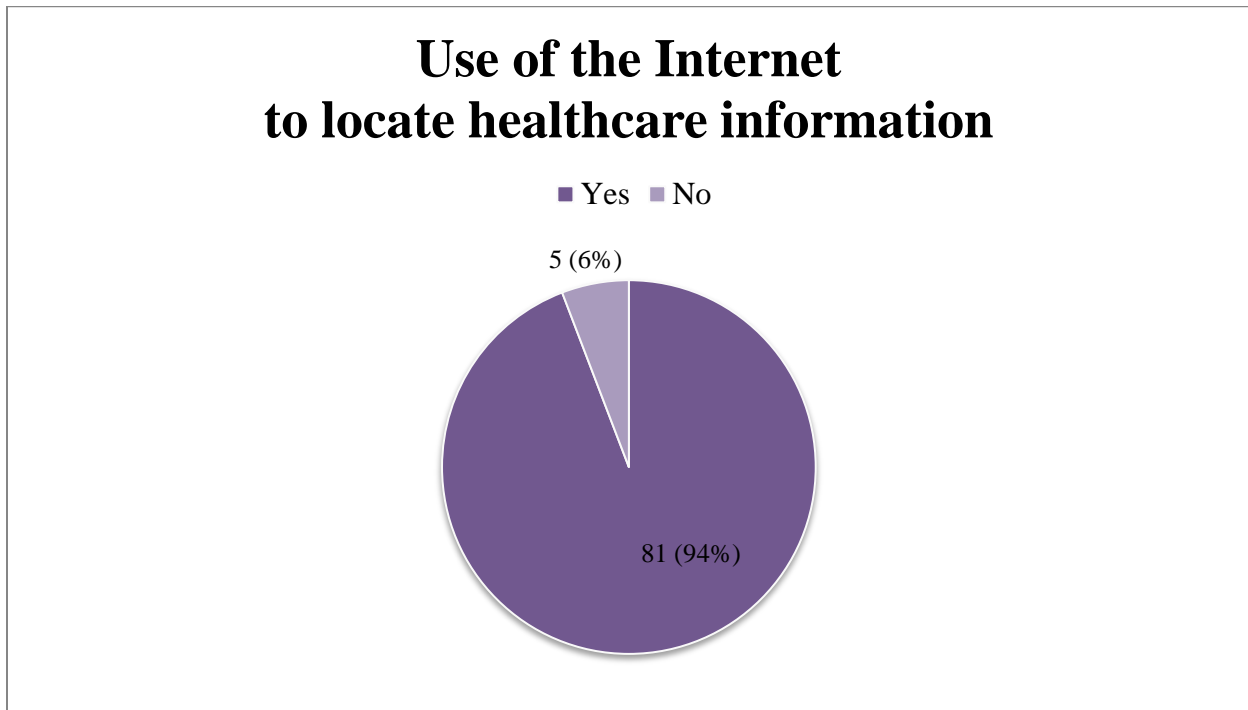


Figure 7. Respondents (not) using the Internet to locate healthcare information

Most of the 81 participants (63%) stated that they consult the Internet to gather information on public health threats, such as health crises (Figure 8). The 30 respondents who reported not searching for information on health crises online mainly provided as a reason the fact that they had never been personally affected by such an event.

**Use of the Internet
to locate information on health crises**
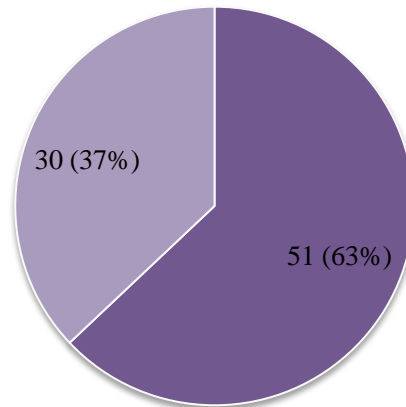
■ Yes  ■ No

30 (37%)

51 (63%)

Figure 8. Respondents (not) using the Internet to locate information on health crises

Regarding the type of health-related content usually sought, answers are reported in Figure 9. Respondents could indicate/select more than one option. The symptoms, treatments and causes of illnesses were the most selected. Findings were broken down by gender. Regardless of gender, symptoms emerged as the type of content that most participants seek online. The second most searched type of healthcare information seems to be prevention of illnesses for men (Figure 10), and treatments of illnesses for women (Figure 11).
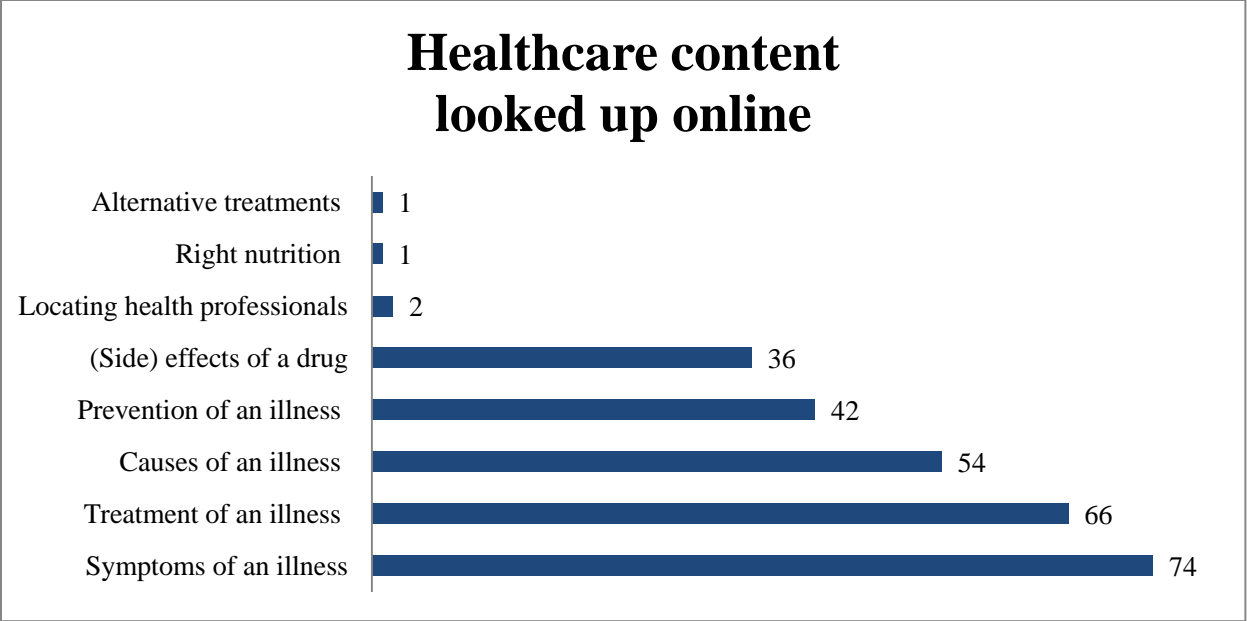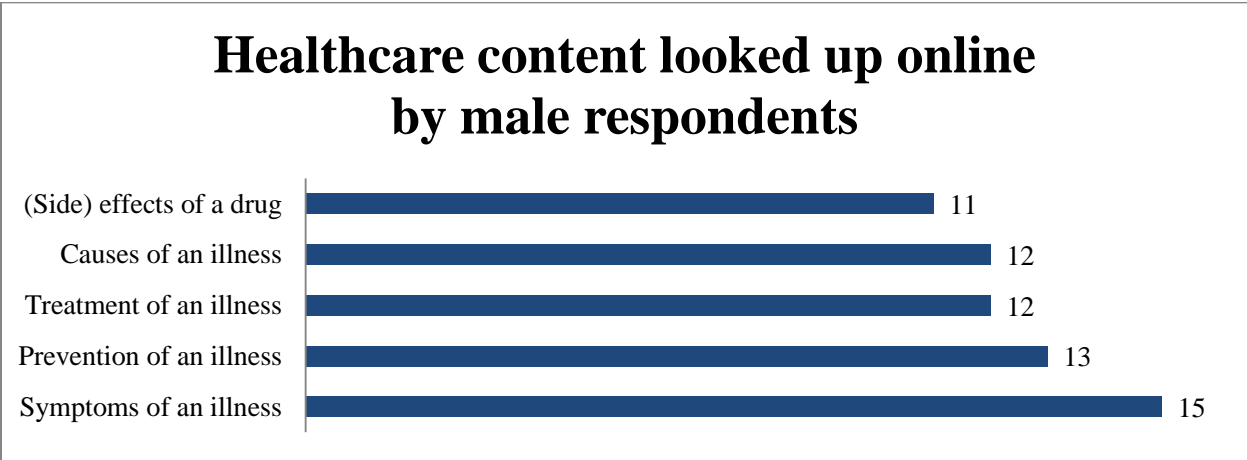
# Healthcare content looked up online

| Category | Value |
|---|---|
| Alternative treatments | 1 |
| Right nutrition | 1 |
| Locating health professionals | 2 |
| (Side) effects of a drug | 36 |
| Prevention of an illness | 42 |
| Causes of an illness | 54 |
| Treatment of an illness | 66 |
| Symptoms of an illness | 74 |

Figure 9. Healthcare information sought online

# Healthcare content looked up online by male respondents

| Category | Value |
|---|---|
| (Side) effects of a drug | 11 |
| Causes of an illness | 12 |
| Treatment of an illness | 12 |
| Prevention of an illness | 13 |
| Symptoms of an illness | 15 |

Figure 10. Healthcare information that male respondents seek online

# Healthcare content looked up online by female respondents

| Category | Value |
|---|---|
| (Side) effects of a drug | 25 |
| Prevention of an illness | 29 |
| Causes of an illness | 41 |
| Treatment of an illness | 53 |
| Symptoms of an illness | 58 |
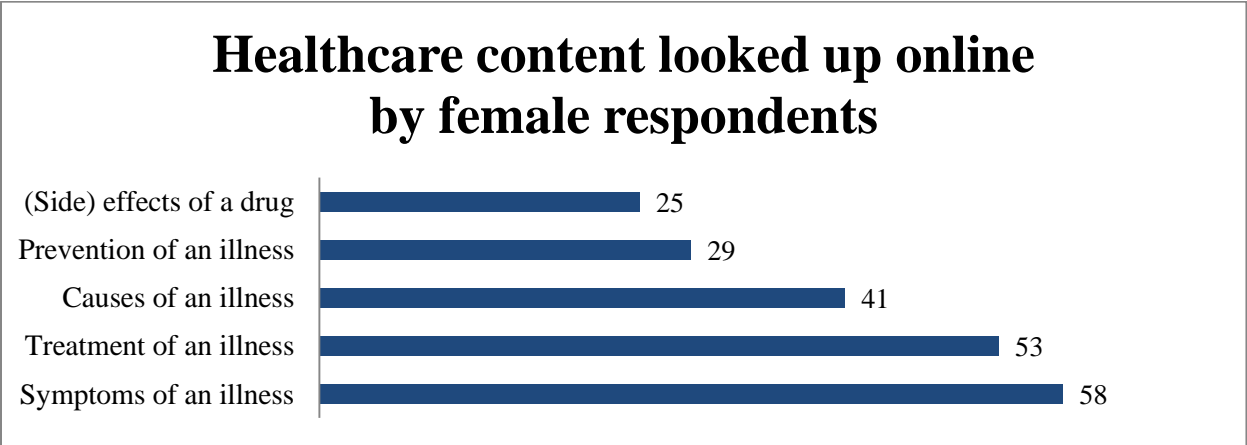
Figure 11. Healthcare information that female respondents seek online

Results on most frequently consulted websites are reported in Figure 12. Respondents could indicate/select more than one option. Wikipedia was the most selected, followed by the WHO website and the website of the Centers for Disease Control and Prevention. The Irish Health Service Executive (HSE) website was mentioned as well.
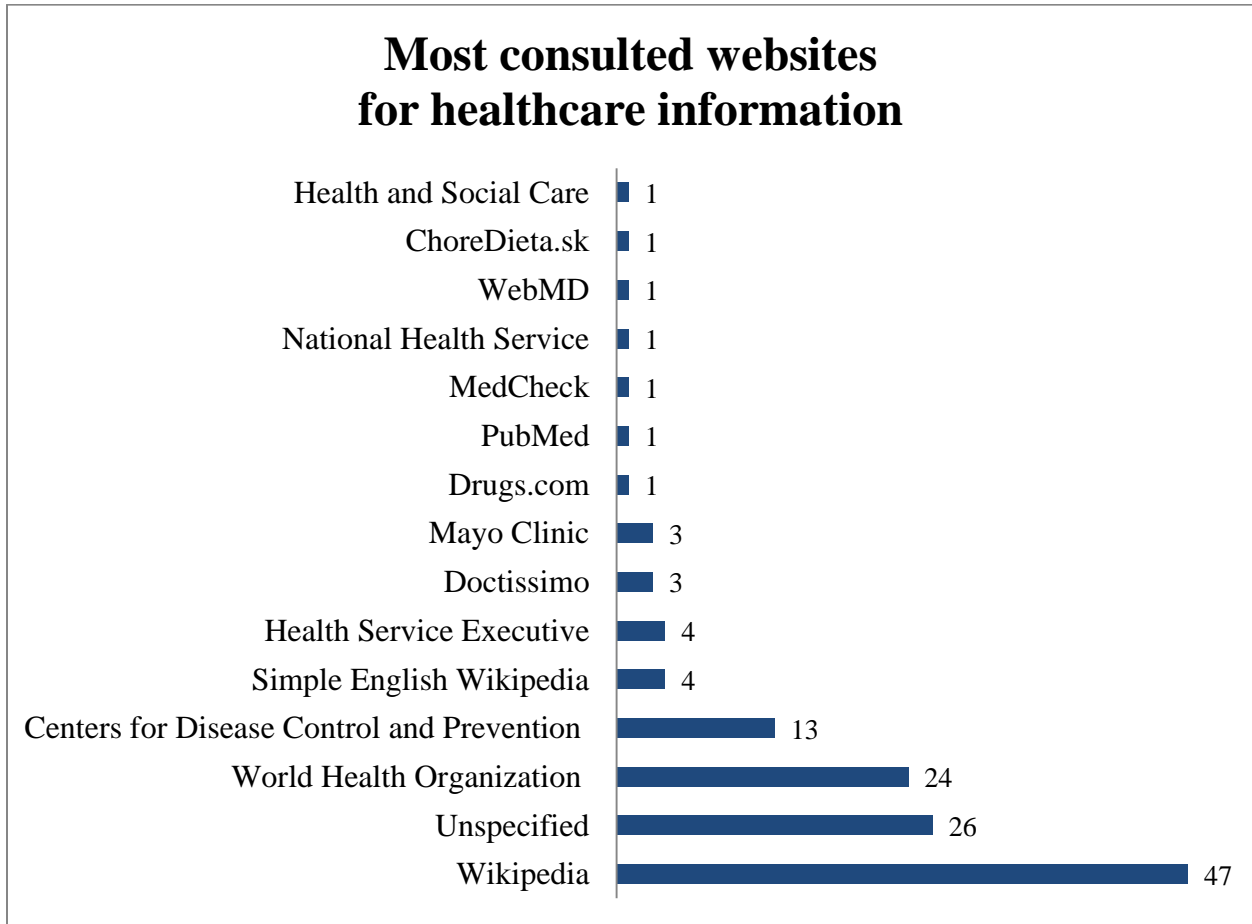


Figure 12. Most consulted websites

### 4.3 Linguistic aspects of searches

Since several websites (e.g. Wikipedia) are available in more than one language, respondents were asked to report the language in which they conduct most of their online health-related searches. One answer could not be interpreted. Figure 13 shows that a small majority of respondents conduct most searches in English. When proficient respondents are separated from LEP respondents, it

emerges that the relative majority of LEP respondents mainly seek healthcare information in their native languages (Figure 14), while the relative majority of proficient respondents conduct most searches in English (Figure 15).
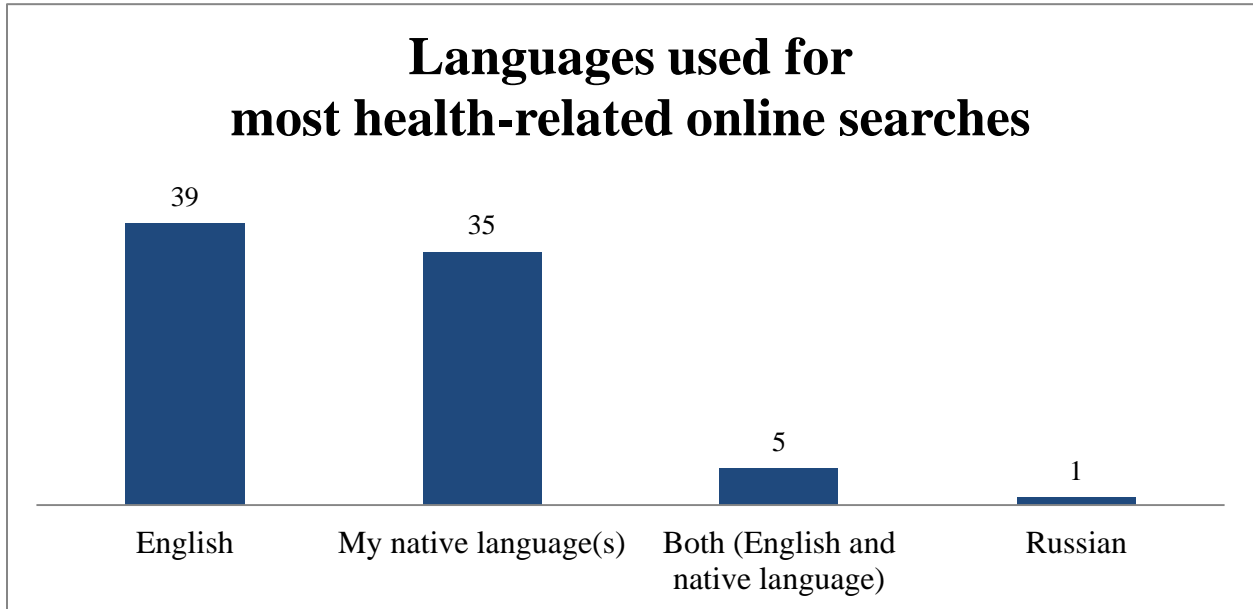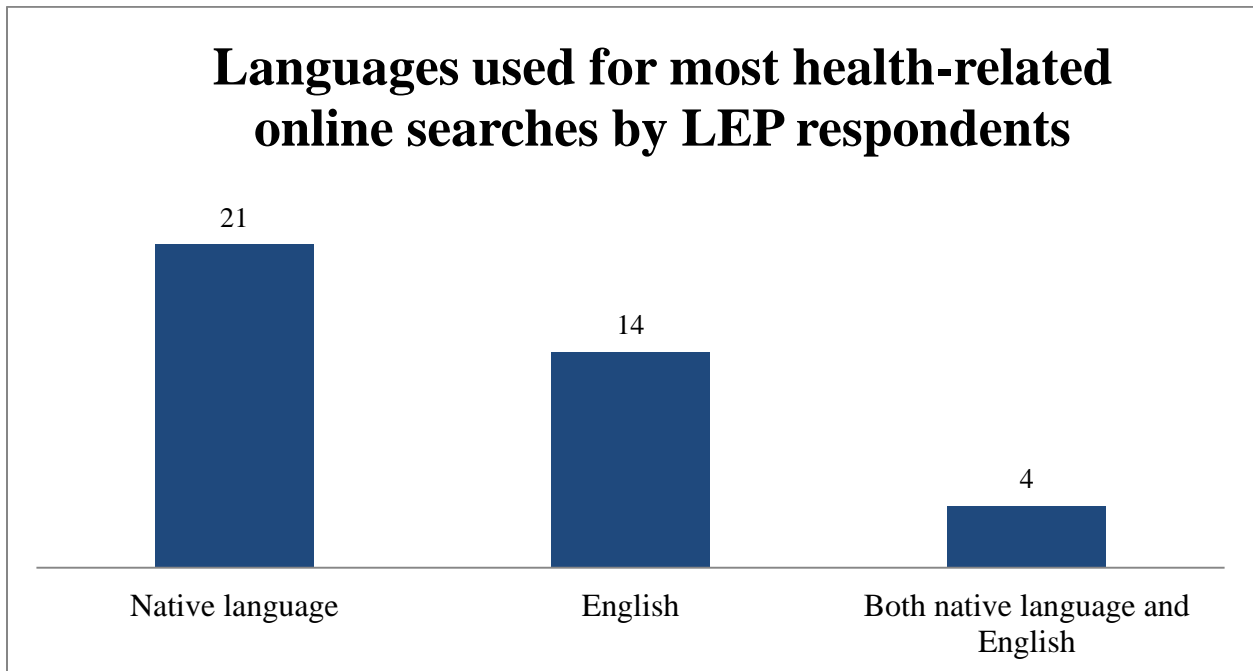
## Languages used for most health-related online searches

English: 39
My native language(s): 35
Both (English and native language): 5
Russian: 1

Figure 13. Languages of most health-related Internet searches

## Languages used for most health-related online searches by LEP respondents

Native language: 21
English: 14
Both native language and English: 4

Figure 14. Languages mostly used by LEP respondents

## Languages used for most health-related online searches by proficient respondents
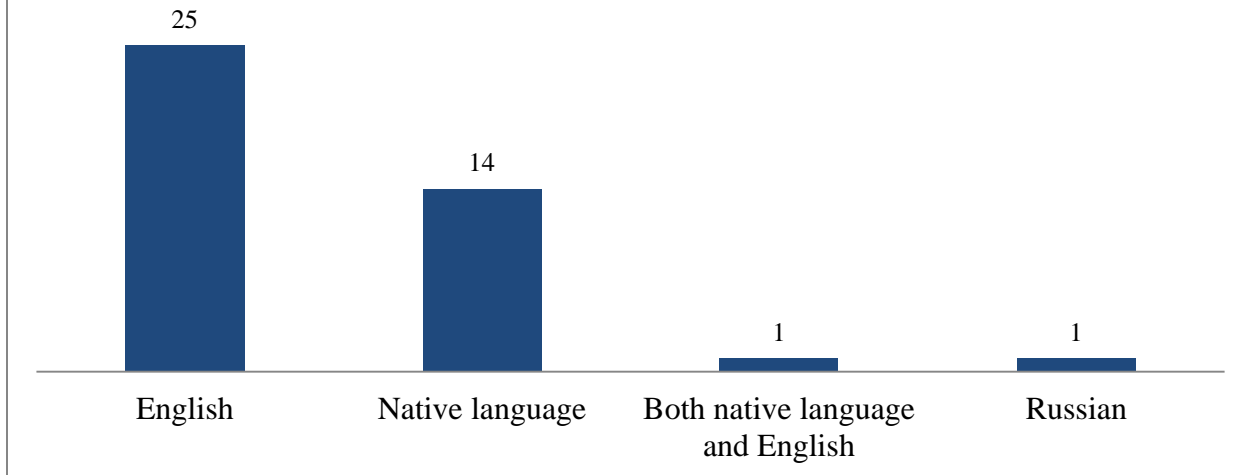
Figure 15. Languages mostly used by proficient respondents

The 39 respondents who reported mainly using English provided different reasons (Figure 16). The main reasons seem to be the higher availability and reliability of English content. For the 35 respondents who reported mainly using their native languages, the main reason seems to be the higher understandability of content (Figure 17). For these questions, participants could select more than one option.

# Reasons for conducting
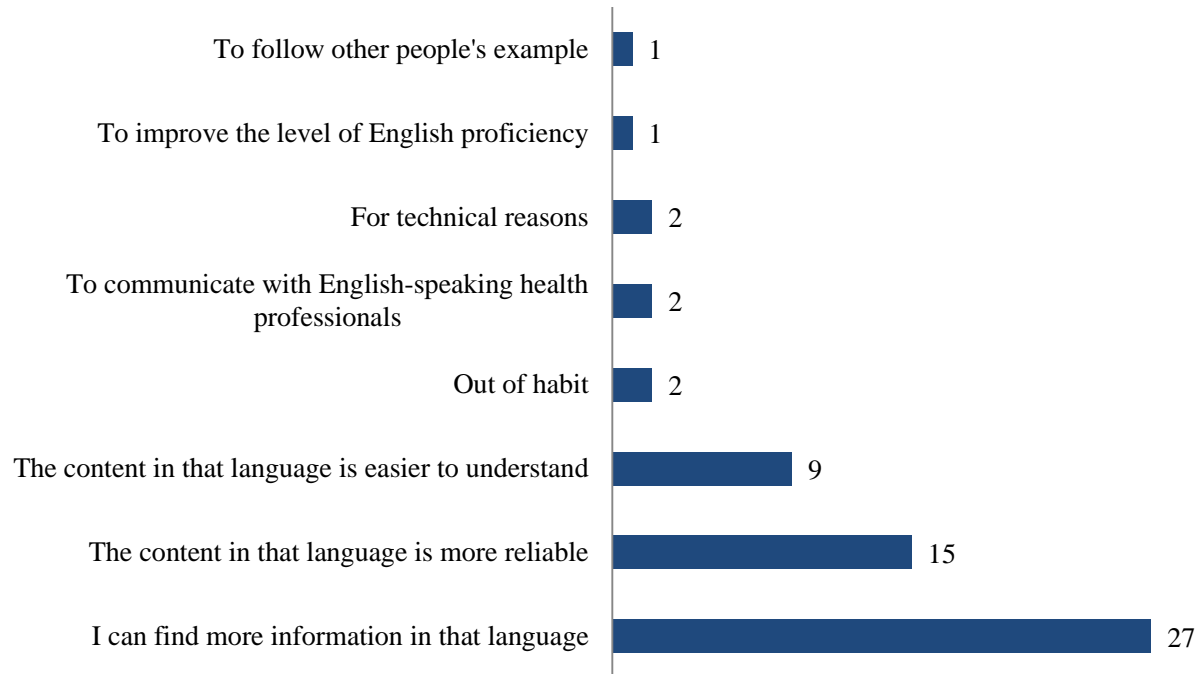# health-related online searches in English



Figure 16. Reasons for conducting searches in English

# Reasons for conducting health-related
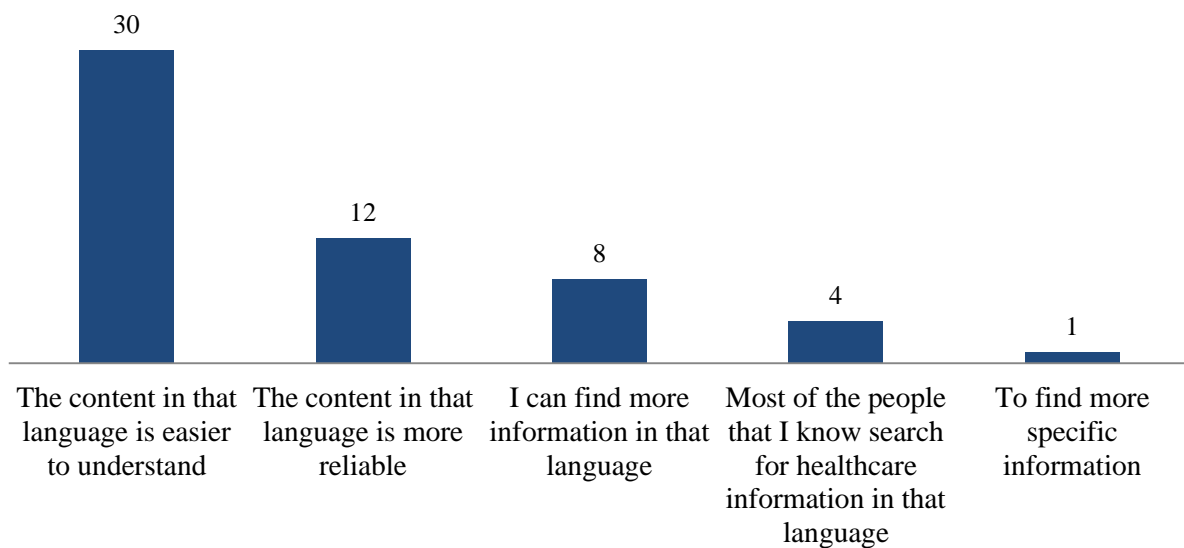# online searches in one's native language



Figure 17. Reasons for conducting searches in one's native language

### 4.4 Comprehension issues

Regarding comprehension issues, results from the participants mainly using English are reported in Figure 18 (one participant skipped this question). Most respondents reported not encountering comprehension problems with English health content. Similarly, among the 35 participants who reported conducting most searches in their native languages, the vast majority (74%) reported not having comprehension issues with health content written in their native languages (Figure 19).
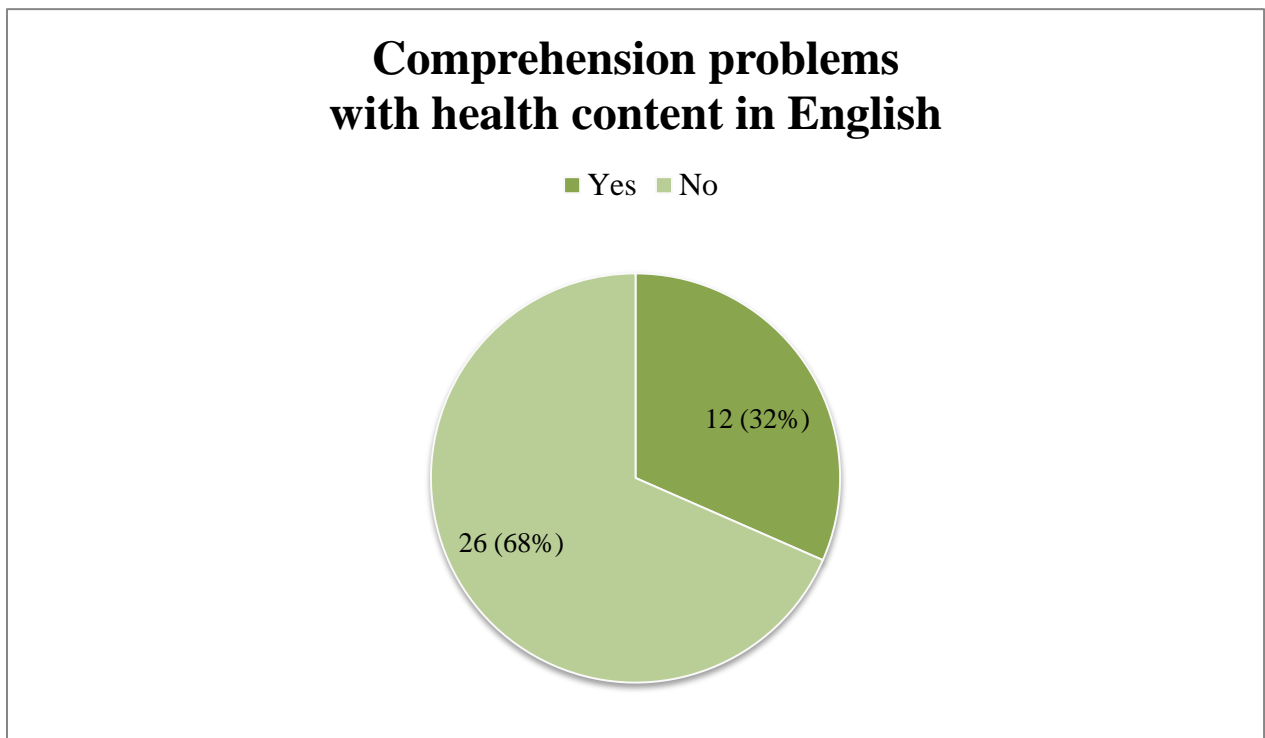
## Comprehension problems
## with health content in English

■ Yes  ■ No

12 (32%)

26 (68%)

Figure 18. Respondents (not) encountering comprehension problems in English

**Comprehension problems with health content in one's native language**
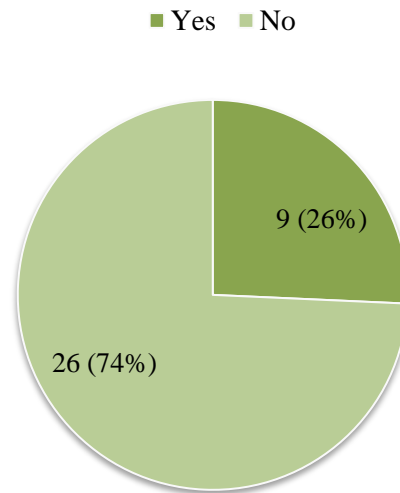
■ Yes ■ No

9 (26%)

26 (74%)

Figure 19. Respondents (not) encountering comprehension problems in their native languages

Respondents who reported comprehension issues were asked to specify which text characteristics hinder their comprehension. They could select/indicate more than one option. 12 participants reported comprehension issues in English, but one skipped this question. All remaining 11 respondents stated that complex medical vocabulary hinders their comprehension. 9 participants reported comprehension issues in their native languages and, again, medical vocabulary received most selections (Figure 20). Complex syntax and low level of cohesion were also mentioned.

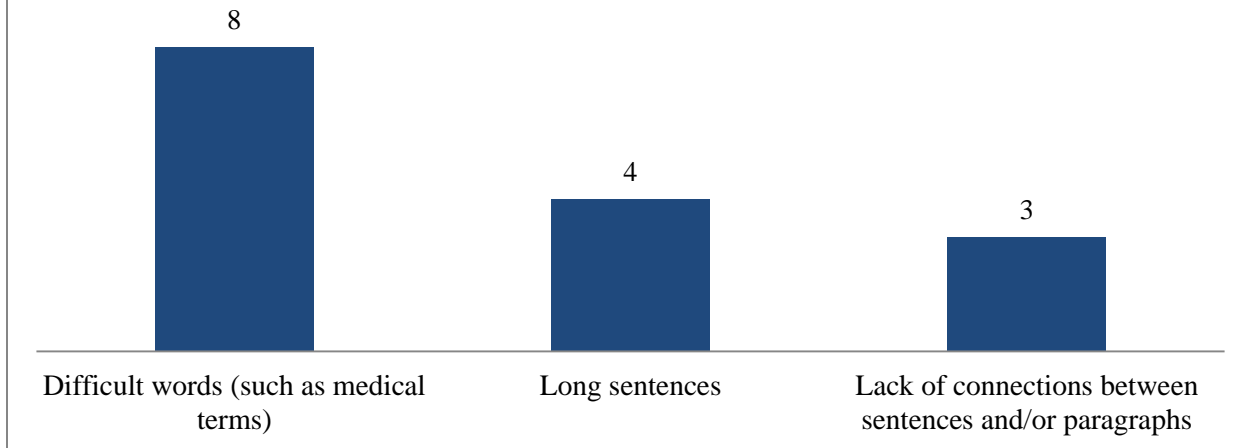**Types of comprehension problems
in one's native language**

Figure 20. Types of comprehension problems in respondents' native languages

### 4.5 Adoption of MT

Figure 21 shows that the majority of the 81 respondents (i.e. 72%) use MT, even though 44% of them use it rarely. More than one quarter of the respondents never use MT systems.

**Frequency of use of MT
for online health content**
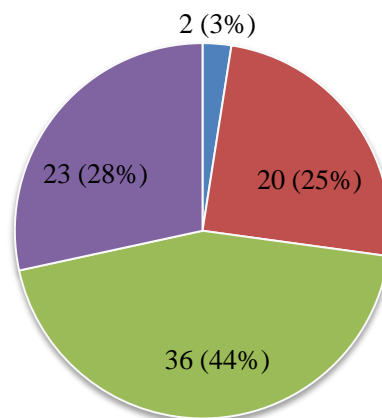
■ Always ■ Often ■ Rarely ■ Never

Figure 21. Frequency of MT use

Findings in Figure 21 were broken down to observe potential differences between LEP respondents and proficient respondents. The percentage of LEP respondents using MT for health content (i.e. 85%) (Figure 22) is higher than the corresponding percentage of proficient respondents (i.e. 59%) (Figure 23). When focusing on the number of participants using MT either always or often, the difference seems more pronounced – 47% of LEP respondents use MT either always or often, against only 9% of proficient respondents.
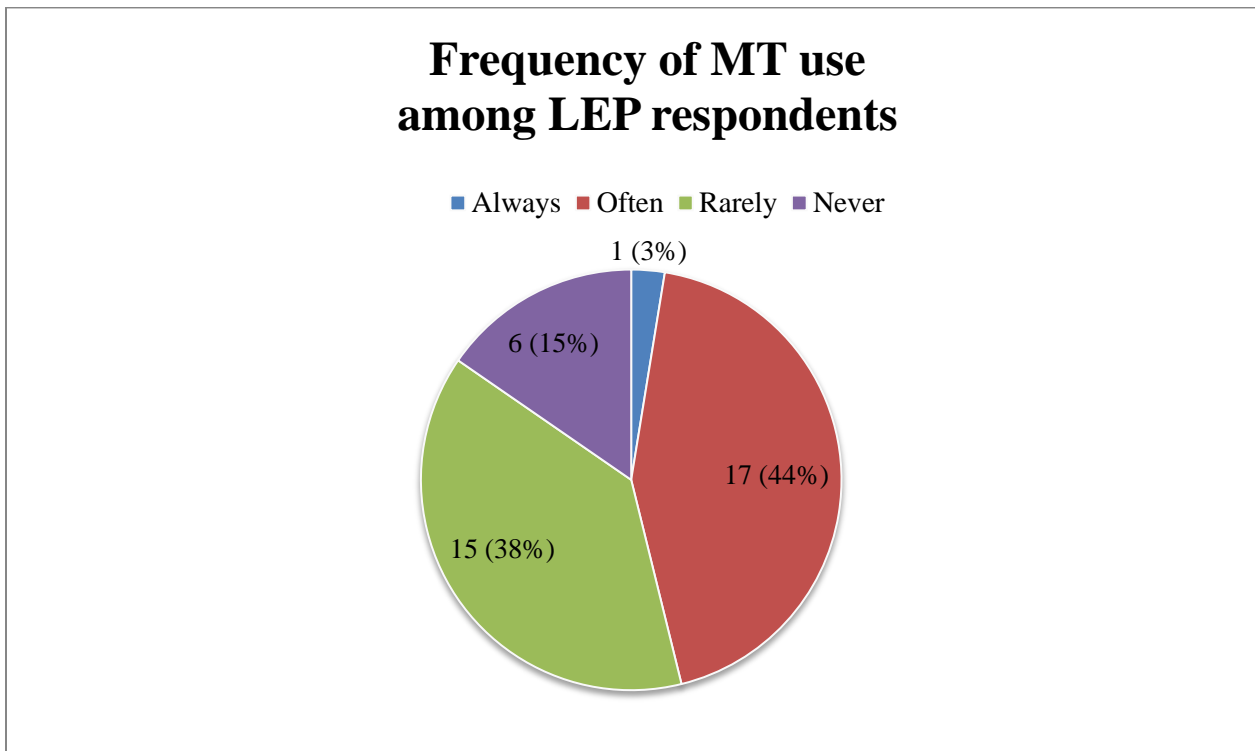


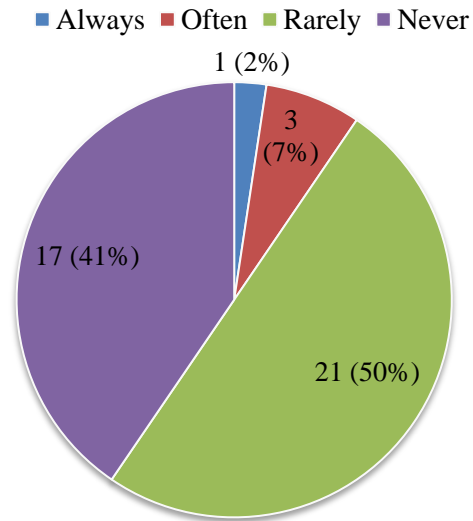Figure 22. Frequency of MT use among LEP respondents

Figure 23. Frequency of MT use among proficient respondents

The 58 respondents who reported using free online MT systems always, often, or rarely were asked to indicate the translation direction(s). Participants could indicate/select more than one option. Figure 24 shows that MT is mainly adopted to translate health content from English into the respondents' native languages.
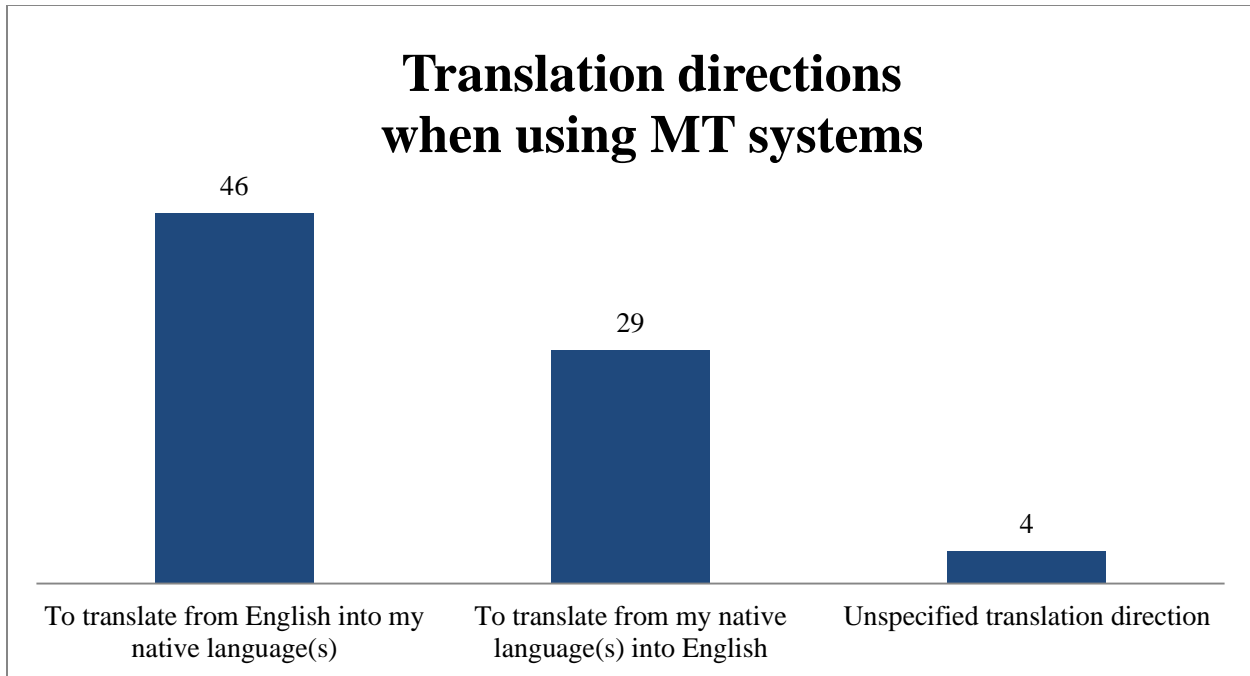
**Translation directions when using MT systems**

46 — To translate from English into my native language(s)

29 — To translate from my native language(s) into English

4 — Unspecified translation direction

Figure 24. Translation directions

## 5. CONCLUSION

This paper described a small-scale survey conducted among non-native speakers of English who had moved to Ireland between 1992 and 2016. The Internet emerged as an important source of healthcare information for users with different language backgrounds. For both men and women, the main focus seems to be on illnesses (their symptoms, treatment, and causes). Moreover, information on health crises seems widely sought online. These findings confirm previous research showing that stressful situations (such as an illness diagnosis or a public health threat) lead people to seek health information online (Weaver et al., 2010). Different websites are used as sources of healthcare information, but Wikipedia received most mentions.

Proficient English speakers seem more likely to conduct their searches in English to benefit from the higher availability of content. In contrast, LEP Internet users show a tendency to search for healthcare information in their native languages, as this facilitates comprehension. This finding shows that having a

limited level of English proficiency might discourage Internet users from accessing vital information and increase their vulnerability, especially if they live in English-speaking countries. For instance, the Irish HSE website provides translations of specific materials (e.g. leaflets). However, the entire website is available in English only.

Text simplification, namely the modification of texts aimed at rendering them more comprehensible (Hill, 1997), might increase the accessibility of online health content for non-native speakers of English. Edits might involve the reduction of lexical and syntactical complexity (Wang et al., 2016), and/or the introduction of cohesive devices (Crossley et al., 2012). Our results indicate that special attention should be given to the simplification of medical vocabulary.

In addition to increasing comprehensibility, simplifying English texts that are then machine translated might improve the quality of the MT output (Mitamura and Nyberg, 1995). MT seems widely adopted among non-native speakers of English living in Ireland, mainly for assimilation purposes. We also found that LEP Internet users adopt MT more frequently than proficient English speakers. It can be assumed these respondents cannot rely on the assistance of a health professional in spotting potential errors/inaccuracies when using free online MT systems.

To conclude, it would be beneficial to conduct a similar survey with a larger sample. It might also be interesting to assess respondents' actual comprehension of health-related texts, and to gather data on additional background characteristics, such as their health literacy levels.

**REFERENCES**

Adams, P., & Fleck, F. (2015). Bridging the language divide in health. *Bulletin of the World Health Organization*, *93*(6), 365-366. doi: 10.2471/BLT.15.020615

Berland, G., Elliott, M., Morales, L., Algazy, J., Kravitz, R., Broder, M., & Watkins, K. (2001). Health information on the Internet: Accessibility, quality, and readability in English and Spanish. *Jama*, *285*(20), 2612-2621.

Crossley, S., Allen, D., & McNamara, D. (2012). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, *16*(1), 89-108. doi: 10.1177/1362168811423456

European Commission (2014). Flash Eurobarometer 404. European citizens' digital health literacy [Report] Retrieved from https://goo.gl/kEkYNB

Gallagher, S., Doherty, D., Moran, R., & Katralova-O'Doherty, K. (2008). *Internet Use and Seeking Health Information Online in Ireland: Demographic Characteristics and Mental Health Characteristics of Users and Non-Users*. Dublin: Health Research Board.

Gan, S. (2012). Lost in translation: How much is translation costing the NHS, and how can we both cut costs and improve service provision? [Report] Retrieved from: https://goo.gl/v2Y2fN

Health Service Executive (2012). Lost in translation? Good practice guidelines for HSE staff in planning, managing and assuring quality translations of health related material into other languages [Report] Retrieved from https://goo.gl/53FzV3

Heilman, J., & West, A. (2015). Wikipedia and medicine: Quantifying readership, editors, and the significance of natural language. *Journal of Medical Internet Research*, *17*(3), e62. doi: 10.2196/jmir.4069

Hill, D. (1997). Survey review: Graded readers. *ELT Journal*, *51*(1), 57-81.

Kirchhoff, K., Turner, A., Axelrod, A., & Saavedra, F. (2011). Application of statistical machine translation to public health information: A feasibility study. *Journal of the American Medical Informatics Association*, *18*(4), 473-478. doi: 10.1136/amiajnl-2011-000176

Mcinnes, N., & Haglund, B. (2011). Readability of online health information: Implications for health literacy. *Informatics for Health and Social Care*, *36*(4), 173-189. doi: 10.3109/17538157.2010.542529

Mitamura, T., & Nyberg, E. (1995). Controlled English for knowledge-based MT: Experience with the KANT system. *Proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation*, 158-172. Retrieved from https://goo.gl/Q1wxSp

Pandya, C., McHugh, M., & Batalova, J. (2011). *Limited English Proficiency Individuals in the United States: Number, Share, Growth, and Linguistic Diversity*. Washington, D.C.: Migration Policy Institute.

Tan, S., & Goonawardene, N. (2017). Internet health information seeking and the patient-physician relationship: A systematic review. *Journal of Medical Internet Research*, *19*(1), e9. doi: 10.2196/jmir.5729

Taylor, R., Crichton, N., Moult, B., & Gibson, F. (2015). A prospective observational study of machine translation software to overcome the challenge of including ethnic diversity in healthcare research. *Nursing Open*, *2*(1), 14-23. doi: 10.1002/nop2.13

Vickstrom, E., Shin, H., Collazo, S., & Bauman, K. (2015). How well - still good? Assessing the validity of the American community survey English-ability question. [Report] Retrieved from https://goo.gl/3pNxy9

Wang, L., Miller, M., Schmitt, M., & Wen, F. (2013). Assessing readability formula differences with written health information materials: Application, results, and recommendations. *Research in Social and Administrative Pharmacy*, *9*(5), 503-516. doi: 10.1016/j.sapharm.2012.05.009

Weaver, J., Mays, D., Weaver, S., Hopkins, G., Eroğlu, D., & Bernhardt, J. (2010). Health information–seeking behaviors, health indicators, and health risks. *American Journal of Public Health*, *100*(8), 1520-1525. doi: 10.2105/AJPH.2009.180521

Wołk, K., & Marasek, K. (2015). Neural-based machine translation for medical text domain. Based on European Medicines Agency leaflet texts. *Procedia Computer Science*, *64*, 2-9. doi: 10.1016/j.procs.2015.08.456

World Health Organisation (2013). Health literacy. The solid facts. [Report] Retrieved from https://goo.gl/81yd6E